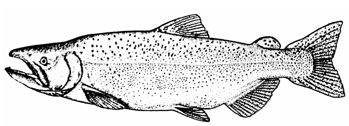
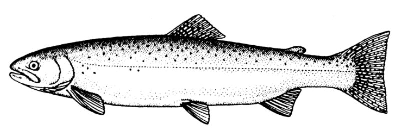
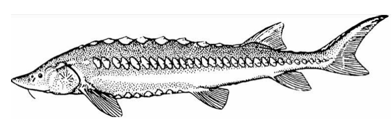
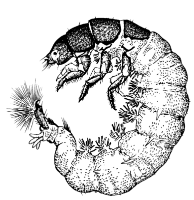
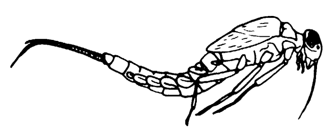
IDENTIFICATION OF THE INSTREAM FLOW REQUIREMENTS FOR ANADROMOUS FISH IN THE STREAMS WITHIN THE CENTRAL VALLEY OF CALIFORNIA AND FISHERIES INVESTIGATIONS

Annual Progress Report
Fiscal Year 2009

U.S. Fish and Wildlife Service
Sacramento Fish and Wildlife Office
2800 Cottage Way, Room W-2605
Sacramento, California 95825

Prepared by staff of
The Energy Planning and Instream Flow Branch

PREFACE

The following is the Eighth Annual Progress Report, Identification of the Instream Flow Requirements for Anadromous Fish in the Streams within the Central Valley of California and Fisheries Investigations, prepared as part of the Central Valley Project Improvement Act (CVPIA) Instream Flow and Fisheries Investigations, an effort which began in October, 2001.[1] Title 34, Section 3406(b)(1)(B) of the Central Valley Project Improvement Act, P.L. 102-575, requires the Secretary of the Department of the Interior to determine instream flow needs for anadromous fish for all Central Valley Project controlled streams and rivers, based on recommendations of the U.S. Fish and Wildlife Service (Service) after consultation with the California Department of Fish and Game (CDFG). The purposes of this investigation are: 1) to provide scientific information to the Service's Central Valley Project Improvement Act Program to be used to develop such recommendations for Central Valley streams and rivers; and 2) to provide scientific information to other CVPIA programs to use in assessing fisheries restoration actions.

The field work described herein was conducted by Ed Ballard, Mark Gard, Bill Pelle, Kevin Aceituno, Jeremy Redding, Rick Williams, Jacob Cunha, Brenda Olson, Tricia Bratcher, Robert Hughes, Steve Thomas and Josh Gruber.

Written comments or questions can be submitted to:

Mark Gard, Senior Biologist
Energy Planning and Instream Flow Branch
U.S. Fish and Wildlife Service
Sacramento Fish and Wildlife Office
2800 Cottage Way, Room W-2605
Sacramento, California 95825

Mark_Gard@fws.gov

[1] The scope of this program was broadened in FY 2009 to include fisheries investigations. This program is a continuation of a 7-year effort, titled the Central Valley Project Improvement Act Instream Flow Investigations, which ran from February 1995 through September 2001.

INTRODUCTION

In response to substantial declines in anadromous fish populations, the Central Valley Project Improvement Act provided for enactment of all reasonable efforts to double sustainable natural production of anadromous fish stocks including the four races of Chinook salmon (fall, late-fall, winter, and spring), steelhead trout, white and green sturgeon, American shad and striped bass. In June 2001, the Service's Sacramento Fish and Wildlife Office, Energy Planning and Instream Flow Branch prepared a study proposal to use the Service's Instream Flow Incremental Methodology (IFIM) to identify the instream flow requirements for anadromous fish in selected streams within the Central Valley of California. The proposal included completing instream flow studies on the Sacramento and Lower American Rivers and Butte Creek which had begun under the previous 7-year effort, and conducting instream flow studies on other rivers, with the Yuba River selected as the next river for studies. The last report for the Lower American River study was completed in February 2003, the final report for the Butte Creek study was completed in September 2003, and the last two reports for the Sacramento River were completed in December 2006. In 2004, Clear Creek was selected as an additional river for studies. In 2007, the Tuolumne River was selected for a minor project to quantify floodplain inundation area as a function of flow. In 2008, South Cow Creek was selected as an additional river for studies. In 2009, the following fisheries investigation tasks were selected for study: 1) Re-examine Clear Creek data on adult Spring Chinook – is the increase in Weighted Useable Area due to an increase in quality or is it an increase in area; 2) Clear Creek Biovalidation – how well does IFIM compare to field observations; 3) Sacramento River tributaries flow and temperature monitoring; 4) Stanislaus River floodplain area versus flow; and 5) Red Bluff Diversion Dam Interim Pumping Plant screen hydraulic evaluation.

The Yuba River study was planned to be a 4-year effort, beginning in September 2001. The goals of the study are to determine the relationship between stream flow and physical habitat availability for all life stages of Chinook salmon (fall- and spring-runs) and steelhead/rainbow trout and to determine the relationship between streamflow and redd dewatering and juvenile stranding. Collection of spawning and juvenile rearing criteria data for fall- and spring-run Chinook salmon and steelhead/rainbow trout was completed by, April 2004 and September 2005, respectively. Field work to determine the relationship between habitat availability for spawning and juvenile rearing and streamflow for spring-run and fall-run Chinook salmon and steelhead/rainbow trout was completed in, FY 2005 and FY 2007, respectively. A draft spawning report was completed in FY 2007 and draft rearing and redd dewatering/juvenile stranding reports were completed in FY 2008. In FY 2008, we completed the response-to-comments document for the peer review of the spawning study report and revisions to the draft spawning study report stemming from the peer review, and conducted a series of stakeholder meetings to discuss stakeholder comments[2] regarding the draft spawning report. In FY 2009, we completed a sensitivity analysis to further respond to concerns raised at those meetings, completed a response-to-comments document for the stakeholder review of the spawning study report and revisions to the draft spawning report stemming from the stakeholder review, and

[2] Stakeholder review for the Yuba reports was agreed upon during scoping meetings prior to commencement of the studies.

conducted a stakeholder review and started a peer review of the juvenile rearing and redd dewatering/juvenile stranding reports. The remaining work on the Yuba reports is ongoing and will be completed in FY 2010.

The Clear Creek study was planned to be a 5-year effort, beginning in October 2003. The goals of the study are to determine the relationship between stream flow and physical habitat availability for all life stages of Chinook salmon (fall- and spring-run) and steelhead/rainbow trout. There are four phases to this study based on the life stages to be studied and the number of segments delineated for Clear Creek from downstream of Whiskeytown Reservoir to the confluence with the Sacramento River[3]. The four phases are: 1) spawning in the upper two segments; 2) fry and juvenile rearing in the upper two segments; 3) spawning in the lower segment; and 4) fry and juvenile rearing in the lower segment. Field work for the above four phases was completed in, FY 2005, FY 2007, FY 2008 and FY 2009, respectively. In FY 2007 the final report and the peer review response-to-comments document for spawning in the upper two segments was completed. A draft report on the five spawning sites in the lower segment was completed in FY 2009 and sent off for stakeholder review. We are currently making arrangements for peer review of that report. In FY 2009, we completed construction of the 2D hydraulic models for four of the five lower segment rearing sites. We are currently awaiting flow data from Graham Matthew and Associates consulting firm needed to calibrate and conduct the production runs for those models. We are also still waiting to receive additional bed topography data on study site 3B in the lower segment from Graham Matthews and Associates. The remaining work on the Clear Creek reports will be completed in FY 2010.

The South Cow Creek study was planned to be a 5-year effort and began in October 2008 with habitat mapping and collection of spawning habitat suitability data for fall-run Chinook salmon. Fieldwork was completed on one site and started on an additional three sites to determine the relationship between stream flow and physical habitat availability for fry and juvenile rearing fall-run Chinook salmon in FY 2009. Due to funding cuts, the South Cow Creek study will be completed in FY 2010 with completion of fieldwork on the three juvenile sites, redd mapping, and preparation of a final report on habitat quantity and quality in South Cow Creek.

For the fisheries investigations tasks, the task “Re-examine Clear Creek data on adult Spring Chinook – is the increase in Weighted Useable Area due to an increase in quality or is it an increase in area” was completed in FY 2009, as was fieldwork for the task “Clear Creek Biovalidation – how well does IFIM compare to field observations.” The latter task will be completed in FY 2010. We began fieldwork for the Sacramento River tributaries flow and temperature monitoring task in FY 2009; due to lack of funding, this task will be continued by the Anadromous Fish Restoration Program Habitat Restoration Coordinators in FY 2010. We were unable to conduct the Stanislaus River floodplain area versus flow task because a U.S.

[3] There are three segments: the upper alluvial segment, the canyon segment, and the lower alluvial segment. Spring-run Chinook salmon spawn in the upper two segments, while fall-run Chinook salmon spawn in the lower segment and steelhead/rainbow trout spawn in all three segments.

Bureau of Reclamation hydraulic model of the Stanislaus River, that would have been used to conduct this task, was not completed in FY 2009. This task will not be conducted in FY 2010 due to lack of funding. In collaboration with the CDFG and the National Marine Fisheries Service, we conducted an initial hydraulic evaluation of the Red Bluff Interim Pumping Plant screens on June 1-12, 2009, and plan to conduct an additional three hydraulic evaluations in FY 2010 at a range of Sacramento River flows and pumping levels.

The following sections summarize project activities between October 2008 and September 2009.

YUBA RIVER

Habitat Simulation

Chinook salmon and steelhead/rainbow trout spawning

A draft report and response to peer review comments document was completed in FY 2007. In FY 2007, we sent out the draft report to interested parties for review and comment prior to finalizing the report. This review by interested parties was in response to commitments made by the Service during the initial planning meetings with those interested parties. This is the first of the CVPIA instream flow reports to be reviewed in this manner. In FY 2008 and 2009, we conducted a series of meetings with stakeholders regarding the draft report. In response to comments received at these meetings, we completed in FY 2009 a habitat modeling and biological verification sensitivity analysis to address these comments. The sensitivity analysis included different methods for developing criteria (density-based criteria), different methods of calculating habitat (geometric mean), and alternative criteria (specifically steelhead/rainbow trout spawning criteria that we developed on Clear Creek). In FY 2009, we completed a response-to-comments document for the stakeholder review of the spawning study report and revisions to the draft spawning report stemming from the stakeholder review. With a second peer review upcoming, a final report on flow-habitat relationships for spawning and the response-to-comments document will be completed in FY 2010.

Juvenile Chinook salmon and steelhead/rainbow trout rearing

Computation of spring/fall-run Chinook salmon and steelhead/rainbow trout fry and juvenile rearing habitat over a range of discharges in was completed for all juvenile rearing sites in FY 2008. The draft report was completed in FY 2008. We sent this draft report out for concurrent stakeholder and peer review in FY 2009. Peer review, response-to-comments document and a final report on flow-habitat relationships for rearing will be completed by September 2010.

Chinook salmon and steelhead/rainbow trout juvenile stranding and redd dewatering

A draft report was completed in FY 2008. We sent this draft report out for concurrent stakeholder and peer review in FY 2009. We will complete the final report in FY 2010.

CLEAR CREEK

Hydraulic and Structural Data Collection

Juvenile fall-run Chinook salmon and steelhead/rainbow trout rearing (Lower Alluvial Segment)

During FY 2008, we completed all the data collection for the Side Channel Run/Pool, North State Riffle, and 3B sites. In FY 2009, we completed the remaining bed topography data collection in the Tarzan Pool site, tied together the vertical benchmarks and collected the 50 validation velocity data points for that site. We then collected the bed topography data, tied together the vertical benchmarks and collected the 50 validation velocity data points for ACID Glide site. We also collected medium and high flow water surface elevations for all five lower rearing sites. Data collection for the five study sites was completed by April 2009. We are still awaiting some additional bed topography data for the 3B study site from Graham Matthews and Associates.

Hydraulic Model Construction and Calibration

Fall-run Chinook salmon and steelhead/rainbow trout spawning (Lower Alluvial Segment)

All data have been compiled and checked, and hydraulic model construction and calibration was completed on all five study sites in FY 2008. We completed the productions runs for all five study sites in early FY 2009.

Fall-run Chinook salmon and steelhead/rainbow trout rearing (Lower Alluvial Segment)

All data collected in FY 2008 for the four study sites has been entered into spreadsheets. We completed hydraulic model construction for four of the five study sites (with the exception of 3B) in FY 2009. The hydraulic model construction for site 3B has been postponed until FY 2010, while we wait for additional bed topography data from Graham Matthews and Associates. We plan on conducting the calibration and production runs for the five study sites in FY 2010 after we receive needed flow data from Graham Matthews and Associates.

Habitat Suitability Criteria (HSC) Development

Juvenile spring-run Chinook salmon and steelhead/rainbow trout rearing (Upper Alluvial and Canyon Segments)

Staff of the Red Bluff Fish and Wildlife Office have been conducting snorkeling surveys specifically to collect rearing HSC for juvenile spring-run Chinook salmon and steelhead/rainbow trout in the Upper Alluvial and Canyon segments. The collection of Young of Year (YOY) spring-run Chinook salmon and steelhead/rainbow trout (fry and juveniles) rearing HSC data began at the end of FY 2004 with surveys conducted on the dates in Table 1. Snorkel surveys were conducted along the banks and in the middle of the channel. Depth, velocity, adjacent velocity[4] and cover data were also collected on locations which were not occupied by YOY spring-run Chinook salmon and steelhead/rainbow trout (unoccupied locations). This was done so that we could apply a method presented in Guay et al. (2000) to explicitly take into account habitat availability in developing HSC criteria, without using preference ratios (use divided by availability). Traditionally, criteria are created from observations of fish use by fitting a nonlinear function to the frequency of habitat use for each variable (depth, velocity, cover, adjacent velocity). One concern with this technique is what effect the availability of habitat has on the observed frequency of habitat use. For example, if cover is relatively rare in a stream, fish will be found primarily not using cover simply because of the rarity of cover, rather than because they are selecting areas without cover. Guay et al. (2000) proposed a modification of the above technique where habitat suitability criteria data are collected both in locations where fish are present and in locations where fish are absent. Criteria are then developed by using a logistic regression with presence or absence of fish as the dependent variable and depth, velocity, cover and adjacent velocity as the independent variables, and all of the data (in both occupied and unoccupied locations) are used in the regression.

Before going out into the field, a data book was prepared with one line for each unoccupied location where depth, velocity, cover and adjacent velocity would be measured. Each line had a distance from the bank, with a range of 0.5 to 10 feet by 0.5 foot increments, with the values produced by a random number generator. In areas that could be sampled up to 20 feet from the bank, the above distances were doubled.

When conducting snorkel surveys adjacent to the bank, one person snorkeled upstream along the bank and placed a weighted, numbered tag at each location where YOY spring-run Chinook salmon or steelhead/rainbow trout were observed. The snorkeler recorded the tag number, the species, the cover code[5] and the number of individuals observed in each 10-20 mm size class on a Poly Vinyl Chloride (PVC) wrist cuff. If one person was snorkeling per habitat unit, the side of the creek to be snorkeled would alternate with each habitat unit and would also include

---

[4] The adjacent velocity was measured within 2 feet on either side of the location where the velocity was the highest. Two feet was selected based on a mechanism of turbulent mixing transporting invertebrate drift from fast-water areas to adjacent slow-water areas where fry and juvenile salmon and steelhead/rainbow trout reside, taking into account that the size of turbulent eddies is approximately one-half of the mean river depth (Terry Waddle, USGS, personal communication), and assuming that the mean depth of Clear Creek is around 4 feet (i.e., 4 feet x ½ = 2 feet). This measurement was taken to provide the option of using an alternative habitat model which considers adjacent velocities in assessing habitat quality. Adjacent velocity can be an important habitat variable as fish, particularly fry and juveniles, frequently reside in slow-water habitats adjacent to faster water where invertebrate drift is conveyed. Both the residence and adjacent velocity variables are important for fish to minimize the energy expenditure/food intake ratio and maintain growth.

[5] If there was no cover elements (as defined in Table 2) within 1 foot horizontally of the fish location, the cover code was 0.1 (no cover).

Table 1
Spring-run Chinook Salmon and Steelhead/Rainbow Trout Juvenile HSC Data Collection

Dates	Average Igo[6] Flows (cfs)
September 24, 2004	213
January 14, 21, and 26-27, 2005	283
February 15, 2005	238
April 6 and 20, 2005	250
May 5, 11-13, 16, 23 and 26, 2005	264
June 7, 10, 13 and 23-24, 2005	198
July 28-29, 2005	154
November 22, 2005	199
December 7-8 and 14-16, 2005	216
January 25-26, 2006	194
February 10, 17 and 23, 2006	272
March 9-10, 15-17, 20-21, 27 and 29, 2006	378
April 6, 20-21, 24 and 26, 2006	333
May 1, 5-6, 9-10, 16-17, 24-25 and 30-31, 2006	262
June 6-7, 2006	136
July 5 and 14, 2006	95
August 8, 2006	89
December 7, 15, 18-20 and 29, 2006	240
January 5, 8, 10, 17-19, 25-26 and 30-31, 2007	217
February 1, 5-7, 13-15, 21 and 27, 2007	261
March 7, 2007	255
April 3, 5, 10, 13, 17 and 26-27, 2007	235
May 1, 11, 15-18 and 23-24, 2007	227
June 7, 19 and 21, 2007	167
July 10, 12 and 19-20, 2007	106
January 16-17 and 30, 2008	253
April 29-30, 2008	224

[6] U.S. Geological Survey Gage Number 11372000 on Clear Creek near Igo, CA.

Table 2
Cover Coding System

Cover Category	Cover Code
No cover	0
Cobble	1
Boulder	2
Fine woody vegetation (< 1" diameter)	3
Fine woody vegetation + overhead	3.7
Branches	4
Branches + overhead	4.7
Log (> 1' diameter)	5
Log + overhead	5.7
Overhead cover (> 2' above substrate)	7
Undercut bank	8
Aquatic vegetation	9
Aquatic vegetation + overhead	9.7
Rip-rap	10

snorkeling the middle portion of some units. As an example, the right bank was snorkeled for one habitat unit, the middle of the next habitat unit was then snorkeled, and then the left bank was snorkeled of the next habitat unit and then the process was repeated.[7] The habitat units were snorkeled working upstream, which is generally the standard for snorkel surveys. In some cases when snorkeling the middle of a habitat unit, the difficulty of snorkeling mid-channel required snorkeling downstream. If three people were going to snorkel each unit, one person snorkeled along each bank working upstream, while the third person snorkeled downstream through the middle of the unit. The distance to be snorkeled was delineated by laying out a tape along the bank as described previously for a distance of 150 feet or 300 feet. The average and maximum distance from the water's edge that was sampled, cover availability in the area sampled (percentage of the area with different cover types) and the length of bank sampled (measured

[7]The Sacramento Fish and Wildlife Office Instream Flow Group designates left and right bank looking upstream.

with a 150 or 300-foot-long tape) was also recorded. When three people were snorkeling, cover percentages were collected by each person snorkeling. After completing each unit, the percentages for each person were combined and averaged. The cover coding system used is shown in Table 2.

A 150 or 300-foot-long tape was put out with one end at the location where the snorkeler finished and the other end where the snorkeler began. Three people went up the tape, one with a stadia rod and data book and the other two with a wading rod and velocity meter. At every 20-foot interval along the tape, the person with the stadia rod measured out the distance from the bank given in the data book. If there was a tag within 3 feet of the location, "tag within 3" was recorded on that line in the data book and the people proceeded to the next 20-foot mark on the tape, using the distance from the bank on the next line. If the location was beyond the sampling distance, based on the information recorded by the snorkeler, "beyond sampling distance" was recorded on that line and the recorder went to the next line at that same location, repeating until reaching a line with a distance from the bank within the sampling distance. If there was no tag within 3 feet of that location, one of the people with the wading rod measured the depth, velocity, adjacent velocity and cover at that location. Depth was recorded to the nearest 0.1 foot and average water column velocity and adjacent velocity were recorded to the nearest 0.01 ft/s. Another individual retrieved the tags, measured the depth and mean water column velocity at the tag location, measured the adjacent velocity for the location, and recorded the data for each tag number. Data taken by the snorkeler and the measurer were correlated at each tag location. For the one-snorkeler surveys, the unoccupied data for the mid-channel snorkel surveys was collected by establishing the distance to be snorkeled by laying out the tape on a bank next to the distance of creek that was to be snorkeled. After snorkeling that distance, the line snorkeled was followed down through the middle of the channel and the randomly selected distance at which the unoccupied data was to be collected was measured out toward the left or right bank, alternating with each 20 foot location along the tape. For the three-snorkeler surveys, unoccupied data was collected for each habitat unit snorkeled in this manner by alternating left and right bank or mid-channel for each habitat unit snorkeled. As an example, for the first habitat unit snorkeled, unoccupied data would be collected along the left bank. At the next unit, data would be collected along the right bank. At the next unit, the data would be collected as described previously using the mid-channel line snorkeled. No HSC snorkel surveys were conducted in FY 2009.

Results

To date, there have been 214 observations of YOY spring-run Chinook salmon, and 566 observations of YOY steelhead/rainbow trout (in this case the use of the term observations indicates when a sighting of one or more fish occurred). An observation can include observations of fry (<60 mm in length) and observations of juveniles (≥60 mm). Of the 214 YOY spring-run Chinook salmon observations, there have been 193 spring-run Chinook salmon observations of <60 mm fish and 34 spring-run Chinook salmon observations of ≥60 mm fish.

Of the 566 YOY steelhead/rainbow trout observations, there have been 279 steelhead/rainbow trout observations of <60 mm fish and 314 steelhead/rainbow trout observations of ≥60 mm fish.

A total of 1,175 mesohabitat units have been surveyed to date. A total of 156,741 feet of near-bank habitat and 33,524 feet of mid-channel habitat have been sampled to date. Table 3 summarizes the number of feet of different mesohabitat types sampled to date and Table 4 summarizes the number of feet of different cover types sampled to date. We have developed two different groups of cover codes based on snorkel surveys we conducted on the Sacramento River: Cover Group 1 (cover codes 4 and 7 and composite [instream+overhead] cover), and Cover Group 0 (all other cover codes). A total of 98,446 feet of Cover Group 0 and 56,029 feet of Cover Group 1 in near-bank habitat, and 32,509 feet of Cover Group 0 and 750 feet of Cover Group 1 in mid-channel habitat, have been sampled to date.

Due to the need to complete all of the Clear Creek reports in FY10, no further YOY and juvenile spring-run Chinook salmon and steelhead/rainbow trout HSC data will be collected.

Habitat Simulation

Juvenile spring-run Chinook salmon and steelhead/rainbow trout rearing (Upper Alluvial and Canyon Segments)

Spring-run Chinook salmon and steelhead/rainbow trout rearing habitat will be computed over a range of discharges for the six spawning sites and six rearing sites in the Upper Alluvial and Canyon segments. Completion of this phase of the study will occur in FY 2010, due to the lack of funds for further snorkeling surveys to collect additional HSC data. Given the small number of observations of juvenile spring-run Chinook salmon gathered to date, it may be necessary to utilize the Clear Creek fall-run Chinook salmon juvenile criteria to be developed, spring-run Chinook salmon juvenile rearing HSC data from another creek or river with characteristics similar to Clear Creek, or conduct transferability tests using Clear Creek fall-run HSC or spring-run rearing HSC from another creek or river. The draft report was partially completed in FY 2009. We will complete draft and final reports on the 2-D modeling of the spring-run Chinook salmon and steelhead/rainbow trout rearing in the Upper Alluvial and Canyon segments in FY 2010. The Red Bluff Fish and Wildlife Office has requested that a draft report be distributed to interested parties for comment in addition to peer review, as is being done with the Yuba River Study reports.

Fall-run Chinook salmon and steelhead/rainbow trout spawning (Lower Alluvial Segment)

We completed the hydraulic model production runs for all five study sites over the range of simulation discharges, computed fall-run Chinook salmon and steelhead/rainbow trout spawning habitat over a range of discharges for the five spawning sites and completed a draft report in FY 2009. A peer review and final report will be completed in FY 2010.

Table 3
Distances Sampled for YOY Spring-run Chinook Salmon and Steelhead/Rainbow Trout HSC Data - Mesohabitat Types

Mesohabitat Type	Near-bank habitat distance sampled (ft)	Mid-channel habitat distance sampled (ft)
Main Channel Glide	4,071	744
Main Channel Pool	66,804	12,993
Main Channel Riffle	31,292	7,011
Main Channel Run	52,065	10,395
Side Channel Glide	0	550
Side Channel Pool	1,180	520
Side Channel Riffle	200	365
Side Channel Run	0	664
Cascade	1,129	282

Table 4
Distances Sampled for YOY Spring-run Chinook Salmon and Steelhead/Rainbow Trout HSC Data - Cover Types

Cover Type	Near-bank habitat distance sampled (ft)	Mid-channel habitat distance sampled (ft)
None	48,623	18,372
Cobble	14,901	8,763
Boulder	7,835	4,558
Fine Woody	48,153	465
Branches	23,518	376
Log	1,700	38
Overhead	1,461	26
Undercut	3,049	73
Aquatic Vegetation	5,115	616
Rip Rap	0	0
Overhead + instream	45,101	611

Fall-run Chinook salmon and steelhead/rainbow trout rearing (Lower Alluvial Segment)

We will complete the hydraulic model production runs for all five study sites over the range of simulation discharges, compute fall-run Chinook salmon and steelhead/rainbow trout rearing habitat over a range of discharges for the five spawning sites and five rearing sites and issue draft and final reports in FY 2010.

SOUTH COW CREEK

Habitat Mapping

Juvenile fall-run Chinook salmon rearing

Mesohabitat mapping of South Cow Creek was conducted October 27-30, 2008, November 24-26, 2008, and April 16, 2009. There were three portions of the creek that were mesohabitat typed. These three sections were the Boero Reach, Valley Floor Reach, and the Tetrick Reach. The combined distance for these three reaches was 7.36 miles. Using habitat typing protocols developed by CDFG, the mesohabitat mapping consisted of walking upstream or downstream and delineating the mesohabitat units. The location of the upstream and downstream boundaries of habitat units was recorded with a Real Time Kinematic (RTK) Global Positioning System (GPS) unit. The mesohabitat units were also delineated on aerial photos.

Following the completion of the mesohabitat mapping on April 16, 2009, the mesohabitat types and number of habitat units of each habitat type in each segment were enumerated, and shapefiles of the mesohabitat units were created in a Geographic Information System (GIS) using the GPS data and aerial photos flown on October 27, 2008. Since we were not able to get permission for access to the upper 1.54 miles of the Valley Floor Reach, identification of habitat types and shapefiles for this area was made solely using the October 27, 2008 aerial photos. The area of each mesohabitat unit was computed in GIS from the above shapefiles. A total of 444 mesohabitat units were mapped for the three reaches. Table 5 summarizes the mesohabitat types, area totals and numbers of each type recorded during the habitat mapping process.

During the course of conducting the mesohabitat mapping, we also attempted to collect fall-run Chinook salmon spawning HSC. We were only able to locate a total of 20 redds, which were insufficient data for use in developing spawning HSC.

Field Reconnaissance and Study Site Selection

Juvenile fall-run Chinook salmon rearing

Field reconnaissance in April and May 2009 investigated potential study sites in the Boero and Valley Floor reaches. Based on the results of the mesohabitat mapping and field reconnaissance, a list of potential study sites was developed. Using the final list of potential study sites, we selected five habitat study sites that will represent the habitat types found in the Boero and

Table 5
FY 2009 South Cow Creek Mesohabitat Mapping Results

Mesohabitat Type	South Cow Creek Units Area Totals (ft^2)	Number of Units
Side Channel Pool	51,292	32
Main Channel Pool	697,366	94
Side Channel Riffle	19,584	40
Main Channel Riffle	261,901	124
Side Channel Run	15,277	13
Main Channel Run	234,679	100
Side Channel Glide	1,156	2
Main Channel Glide	138,234	37
Cascade	493	2

Valley Floor reaches. We randomly selected the habitat study sites to insure unbiased selection of the study sites. The following is the final list of the five study sites, listed in order from upstream to downstream: Jones, Poole, Farrell, Sabanovich and Boero.

Transect Placement (study site setup)

Juvenile fall-run Chinook salmon rearing

The Poole, Jones, Sabanovich, and Farrell study sites were established in April 2009, while the Boero study site was established in May 2009. For the sites selected for modeling, the landowners along both riverbanks were identified and temporary entry permits were sent, accompanied by a cover letter, to acquire permission for entry onto their property during the course of the study.

For each study site, a transect was placed at the up- and downstream ends of the site. The downstream transect will be modeled with the Physical Habitat Simulation System (PHABSIM) to provide water surface elevations as an input to the 2-D model. The upstream transect will be used in calibrating the 2-D model. The initial bed roughnesses used by River2D are based on the observed substrate sizes and cover types. A multiplier is applied to the resulting bed roughnesses, with the value of the multiplier adjusted so that the WSEL generated by River2D at the upstream end of the site match the WSEL predicted by the PHABSIM transect at the upstream end of the site. Transect pins (headpins and tailpins) were marked on each river bank above the 300 cfs water surface level using rebar driven into the ground and/or bolts placed in tree trunks. Survey flagging was used to mark the locations of each pin. We also installed horizontal bench marks that act as control points for the bed topography data collection when using a robotic total station. After installing the horizontal bench marks, data was collected to establish a precise set of location coordinates for each horizontal bench mark using survey-grade RTK GPS. Vertical benchmarks (lagbolts in trees or bedrock points) were established, and marked with paint and flagging.

Hydraulic and Structural Data Collection

Juvenile fall-run Chinook salmon rearing

Hydraulic and structural data collection for the Boero study site was completed in FY 2009. Low and medium flow water surface elevations were collected for all five sites. Velocity sets were collected for the transects at the Boero, Poole, Jones, and Farrell sites. Depth and velocity measurements were made by wading with a wading rod equipped with a Marsh-McBirney[R] model 2000 or a Price AA velocity meter. A tape was used to measure stations along the transects. Substrate and cover (Tables 6 and 2) along the transects were determined visually. Dry bed elevations and substrate and cover data along the transects were collected and the vertical benchmarks were tied together for the Boero, Poole and Jones sites. Due to lack of sufficient funds and time constraints, we were unable to collect data on the Sabanovich study site and eliminated it from the study.

We collected the data between the inflow and outflow transects by obtaining the bed elevation and horizontal location of individual points with a total station or survey-grade RTK GPS, while the cover and substrate was visually assessed at each point. Bed topography data collection was completed for the Boero study site and a majority of the data was collected for the Poole, Jones, and Farrell sites. Stage of zero flow at the outflow transect was surveyed in for the Boero, Poole, and Jones sites. We anticipate collecting high flow water surface elevations during the winter of 2009-2010 on the four study sites. We will also complete the bed topography data collection on the Poole, Jones, and Farrell study sites in FY 2010.

To validate the velocities predicted by the 2-D model within the Boero, Poole, and Jones study sites, depth, velocities, substrate and cover measurements were collected in the site by wading with a wading rod equipped with a Marsh-McBirney model 2000 velocity meter. The horizontal locations and bed elevations were determined by taking a total station shot on a prism held at each point where depth and velocity were measured for these sites. A total of 50 representative points were measured throughout each site. We anticipate completing the hydraulic and structural data collection for the four rearing sites in FY 2010.

Table 6
Substrate Descriptors and Codes

Code	Type	Particle Size (inches)
0.1	Sand/Silt	< 0.1
1	Small Gravel	0.1 – 1
1.2	Medium Gravel	1 – 2
1.3	Medium/Large Gravel	1 – 3
2.3	Large Gravel	2 – 3
2.4	Gravel/Cobble	2 – 4
3.4	Small Cobble	3 – 4
3.5	Small Cobble	3 – 5
4.6	Medium Cobble	4 – 6
6.8	Large Cobble	6 – 8
8	Large Cobble	8 – 10
9	Boulder/Bedrock	> 12
10	Large Cobble	10 – 12

Hydraulic Model Construction and Calibration

Juvenile fall-run Chinook rearing

The topographic data for the 2-D model (contained in bed files) is first processed using the R2D_Bed software, where breaklines are added to produce a smooth bed topography. The resulting data set is then converted into a computational mesh using the R2D_Mesh software, with mesh elements sized to reduce the error in bed elevations resulting from the mesh-generating process to 0.1 foot where possible, given the computational constraints on the number of nodes. The resulting mesh is used in River2D to simulate depths and velocities at the flows to be simulated.

The PHABSIM transect at the outflow end of each site is calibrated to provide the WSEL at the outflow end of the site used by River2D. The PHABSIM transect at the inflow end of the site is calibrated to provide the water surface elevations used to calibrate the River2D model. The initial bed roughnesses used by River2D are based on the observed substrate sizes and cover types. A multiplier is applied to the resulting bed roughnesses, with the value of the multiplier adjusted so that the WSEL generated by River2D at the inflow end of the site match the WSEL predicted by the PHABSIM transect at the inflow end of the site[8]. The River2D model is run at the flows at which the validation data set was collected, with the output used to determine the difference between simulated and measured velocities, depths, bed elevations, substrate and cover. The River2D model is also run at the simulation flows to use in computing habitat. All data for the four fall-run Chinook salmon rearing sites have been compiled and checked. PHABSIM calibration has been completed for two sites (Boero and Poole sites). Construction and calibration of the 2-D hydraulic model has been completed for the Boero site. Construction and calibration of the 2-D hydraulic models as described above for the three other study sites and running the production runs for the simulation flows for all four sites will be completed in FY 2010.

Habitat Suitability Criteria Development

Juvenile fall-run Chinook salmon rearing

We will be using habitat suitability criteria developed for the Lower Alluvial Segment of Clear Creek for fall-run fry and juvenile Chinook salmon rearing.

Habitat Simulation

Juvenile fall-run Chinook salmon rearing

Using the fall-run Chinook salmon fry and juvenile rearing HSC developed for the Lower Alluvial Segment of Clear Creek, fall-run Chinook salmon fry and juvenile rearing habitat will be computed over a range of discharges for the four rearing sites in South Cow Creek. Completion of this phase of the study will occur in FY 2010. We anticipate completing draft and final reports on the 2-D modeling of the fall-run Chinook salmon juvenile rearing in South Cow Creek in FY 2010.

FISHERIES INVESTIGATIONS

Re-examine Clear Creek data on Adult Spring Chinook

Methods

The purpose of this task was to determine if the increase in Weighted Useable Area (WUA) was due to an increase in habitat quality or due to an increase in area. To accomplish this task, we needed to compute, over a range of flows, the amount of area of spring-run Chinook salmon spawning habitat, to compare it to the amount of WUA from U.S. Fish and Wildlife Service (2007). The amount of area of habitat can be computed in River2D by using binary criteria,

---

[8] This is the primary technique used to calibrate the River2D model.

which have a suitability of 1 for habitat and zero for non-habitat. We developed the binary criteria based on the continuous criteria in U.S. Fish and Wildlife Service (2007) by defining habitat as depths, velocities or substrate categories that had a continuous suitability greater than 0.2 in U.S. Fish and Wildlife Service (2007). The binary criteria were used with the final Computational Mesh file (cdg) production files and the substrate file for each site to compute the area of spring-run Chinook salmon spawning habitat over the desired range of simulation flows for all sites. The area values for the sites in each segment were added together and multiplied by the ratio of total redds counted in the segment to number of redds in the modeling sites for that segment to produce the total area per segment. The spring-run Chinook salmon multipliers were calculated using redd counts from 2000-2005.

Results

The binary criteria that were used to compute the area of spring-run Chinook salmon spawning habitat are given in Table 7, while the comparison of area of spring-run Chinook salmon spawning habitat to the amount of WUA from U.S. Fish and Wildlife Service (2007) is shown in Figures 1 and 2.

Discussion

For the upper alluvial segment, which contains the vast majority of the spring-run Chinook salmon spawning habitat, as compared to the canyon segment, the increase in WUA was due to an increase in area up to a flow of approximately 400 cfs, while the increase in WUA going from 400 to 900 cfs was due to an increase in quality. Specifically, the amount of area of spring-run Chinook salmon spawning habitat changed in the same manner as the amount of WUA from U.S. Fish and Wildlife Service (2007) for flows up to 400 cfs, while the flow-habitat relationship patterns for area and WUA deviated for flows greater than 400 cfs.

Table 7
Binary Criteria Used to Compute Area For Spring-run Chinook Salmon Spawning Habitat

Velocity (ft/s)	**Velocity Suitability**	**Depth (ft)**	**Depth Suitability**	**Substrate Code**	**Substrate Suitability**
0.00	0	0.0	0	0	0
1.19	0	1.1	0	1.2	0
1.20	1	1.2	1	1.3	1
4.40	1	7.0	1	3.4	1
4.41	0	7.1	0	3.5	0
100	0	100	0	100	0

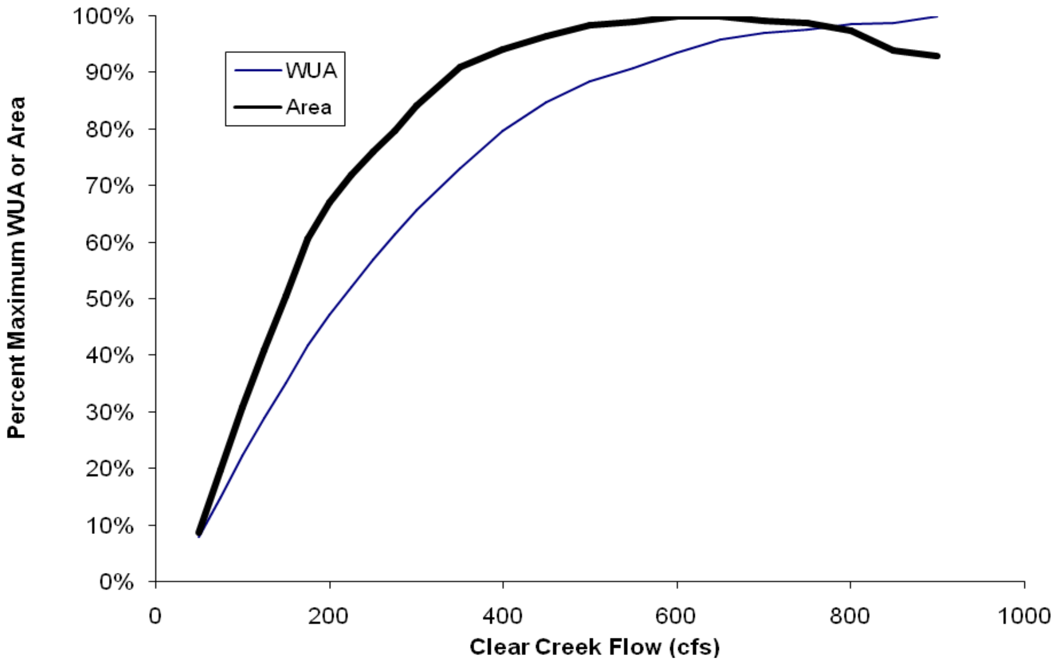


Figure 1
Area and WUA for Spring-run Chinook Salmon Spawning in the Upper Alluvial Reach

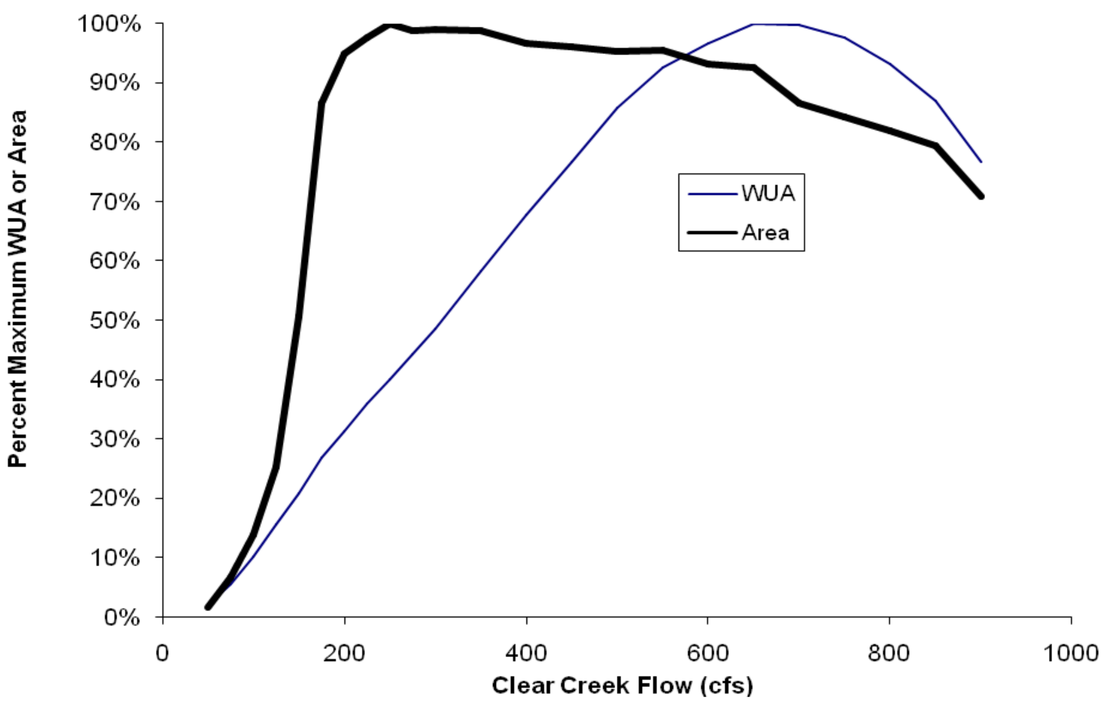


Figure 2
Area and WUA for Spring-run Chinook Salmon Spawning in the Canyon Reach

Clear Creek Biovalidation

Methods

This task had the following six subtasks: 1) compare 2008 juvenile habitat use to juvenile Combined Suitability Index (CSI); 2) compare 2005 juvenile habitat use to juvenile CSI; 3) compare 2007 Spawning Area Mapping (SAM) to adult CSI; 4) compare 2008 SAM to adult CSI; 5) after building fall-run Chinook salmon adult criteria from unoccupieds in model, rerun earlier analysis comparing SAM and CSI; and 6) review statistical approach for these. The juvenile habitat use and spawning area mapping data was supplied by the Red Bluff Fish and Wildlife Office. Discussions during FY 2009 narrowed the scope of this work to examining data from restoration sites 3A and 3B. CSI values for site 3B will be computed from the River2D model developed for the Clear Creek IFIM study. CSI values for site 3A will be computed from a River2D model that will be developed using: 1) bed topography data previously collected by Graham Matthews and Associates; 2) substrate and cover polygon mapping that the Energy Planning and Instream Flow Branch conducted in FY 2009; and 3) transect data collected by the Energy Planning and Instream Flow Branch in FY 2009.

Results

A transect was placed at the up- and downstream ends of the 3A study site. The downstream transect will be modeled with PHABSIM to provide water surface elevations as an input to the 2-D model. The upstream transect will be used in calibrating the 2-D model. The initial bed roughnesses used by River2D are based on the observed substrate sizes and cover types. A multiplier is applied to the resulting bed roughnesses, with the value of the multiplier adjusted so that the WSEL generated by River2D at the upstream end of the site match the WSEL predicted by the PHABSIM transect at the upstream end of the site. Transect pins (headpins and tailpins) were marked on each river bank above the 900 cfs water surface level using rebar driven into the ground and/or bolts placed in tree trunks. Survey flagging was used to mark the locations of each pin. Vertical benchmarks (lagbolts in trees or bedrock points) were established, and marked with paint and flagging. The location coordinates for each transect pin and elevations of the vertical benchmarks were determined using survey-grade RTK GPS.

Low, medium and high flow water surface elevations, dry bed elevations, substrate and cover data, and velocity sets were collected for the transects at the 3A study site in FY 2009. Depth and velocity measurements were made by wading with a wading rod equipped with a Marsh-McBirney^R model 2000 or a Price AA velocity meter. A tape was used to measure stations along the transects. Substrate and cover (Tables 6 and 2) along the transects were determined visually.

Substrate and cover polygons were mapped throughout the 3A study site up to the 900 cfs water surface level using survey-grade RTK GPS in FY 2009. This data will allow us to assign substrate, cover and bed roughness values to each of the bed topography data points previously collected by Graham Matthews and Associates. We plan to conduct hydraulic modeling construction and calibration and habitat simulation for the 3A study site in FY 2010 once we have obtained the bed topography data previously collected by Graham Matthews and Associates. After we have completed the hydraulic modeling construction and calibration and habitat simulation for the 3A and 3B study sites, we will be able to complete the first five subtasks. The sixth subtask was completed in FY 2009 (Appendix A) by Western Ecosystems Technology, Inc. under a Cooperative Agreement funded by the Energy Planning and Instream Flow Branch. We plan to complete this entire task in FY 2010.

Sacramento River Tributaries Flow and Temperature Monitoring

Methods

The purpose of this task was to produce regression formulas that could be used to predict flows and water temperatures for the following tributaries of the Sacramento River using flow and air temperature data available on the Internet: Antelope Creek, South Fork Cottonwood Creek, Stillwater Creek, Churn Creek and Bear Creek. The first step for this task was to identify historical gage flow records that could be used to develop regression formulas to predict flows. Additional flow data was collected in FY 2009 to corroborate the flow/flow regression equations and to develop flow/flow regression equations for tributaries or locations which had never been gaged. Flow measurements were made using a tape to measure stations and by wading with a wading rod equipped with a Marsh-McBirney^R model 2000 or a Price AA velocity meter to measure depths and velocities. Depths were measured to the nearest 0.05 foot and velocities were measured to the nearest 0.01 foot/sec for 20 seconds at 0.6 of the depth. Starting in May, we also noted the presence or absence of flow for three tributaries of Antelope Creek (Butler Slough, Craig Creek and New Creek) at the same time that we measured flows on Antelope Creek. In addition, we deployed HOBO Water Temperature Pro V2 probes, manufactured by Onset Corporation, at the locations where we collected flow data. We installed two probes for each stream for redundancy in case probes were lost due to theft or high flows. Each probe was placed in a 2-inch diameter PVC housing (with holes drilled into it) and caps and secured to trees or other immovable objects near the water's edge with 1/8-inch cable. The thermographs were set up to record water temperatures every half-hour. Thermographs were initially deployed on March 16 and 19, 2009 and data was downloaded from the thermographs every other month with an optical shuttle. Daily average water temperatures were calculated for each thermograph, and then the daily average water temperatures for the two thermographs at each site were averaged to produce the daily average water temperature at each site for each day that data was collected. We used the data we collected to develop flow/flow regression equations for tributaries or locations which had never been gaged and regressions of water temperature versus air temperature and flow.

Results

Table 8 summarizes the historical gage flow records used to develop regression formulas to predict flows, while Table 9 presents the regression formulas. Figures 3 to 6 show the historical gage flows and regression equations. No historical gage data is available for Stillwater Creek.

Table 8
Historical Gage Data Used to Develop Flow/Flow Regresssions

Stream	USGS Gage Number	Period of Record
Cow Creek	11374000	10/1/49-present
Cottonwood Creek	11376000	10/1/40-present
Deer Creek	11383500	10/1/40-present
Churn Creek	11372050	10/1/60-9/30/66
Bear Creek	11374100	10/1/59-9/30/67
South Fork Cottonwood Creek	11375900	6/23/82-9/30/85
Antelope Creek	11379000	10/1/40-9/30/82

Table 9
Flow/Flow Regresssions

Regression Equation	R^2
Churn Creek Flow = Max (0, -4.19 + 0.035 x Cow Creek Flow)	0.565
Bear Creek Flow = Max(4, $10^{(-0.0828 + 0.724 \times \log(\text{Cow Creek Flow}))}$)	0.908
South Fork Cottonwood Creek Flow = Max (0, -59.49 + 0.397 x Cottonwood Creek Flow)	0.959
Antelope Creek Flow = Max(0, -20.4 + 0.4977 x Deer Creek Flow)	0.853

Figures 7 to 10 show the annual average hydrographs for Churn, Bear, South Fork Cottonwood and Antelope Creeks, computed from the period of record flows for Cow, Cottonwood and Deer Creeks and the flow/flow regression equations in Table 9.

We were unable to get access to the location of the historical gage on Antelope Creek since the location is on private land. This gage was located upstream of the Edwards/Los Molinos Mutual diversion dam. The only locations where we were able to get access were downstream of this dam, and thus we needed to develop a new flow/flow regression for Antelope Creek downstream of the Edwards/Los Molinos Mutual diversion dam. In March and May, we measured the flow of Antelope Creek at Highway 99. However, since the flow in May was greater than the flow in March, due to considerable flow coming from a tributary (likely an agricultural return flow) located approximately 75 feet upstream of Highway 99, we moved the discharge location for Antelope Creek to the upper end of Cone Grove Park; discharges were measured at this location in June through September. Based on the thermograph data, it appeared that the flow for Stillwater Creek dropped to zero on July 16, 2009. Flow in Cottonwood Creek had reached zero by September 15, 2009.

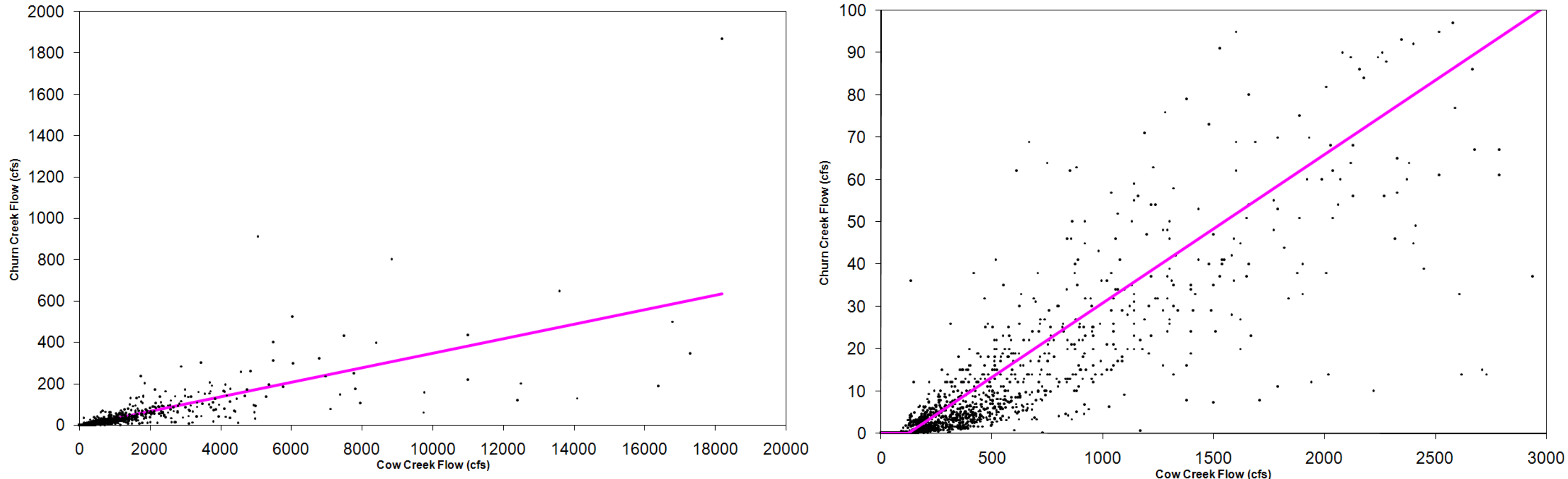


Figure 3
Churn Creek Flow Data and Regression

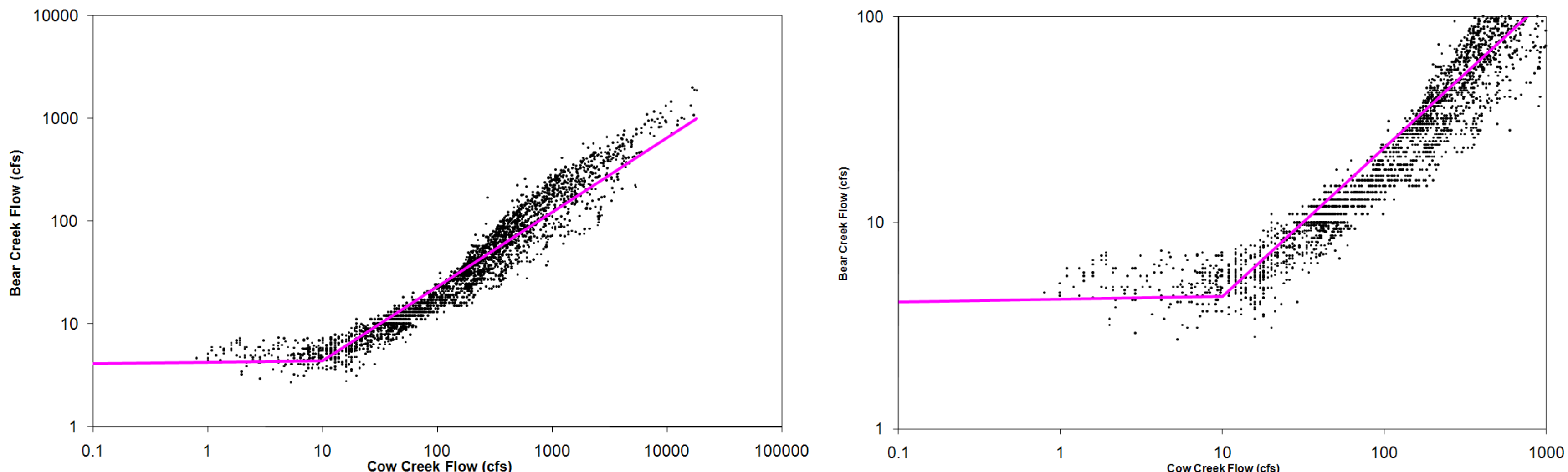


Figure 4
Bear Creek Flow Data and Regression

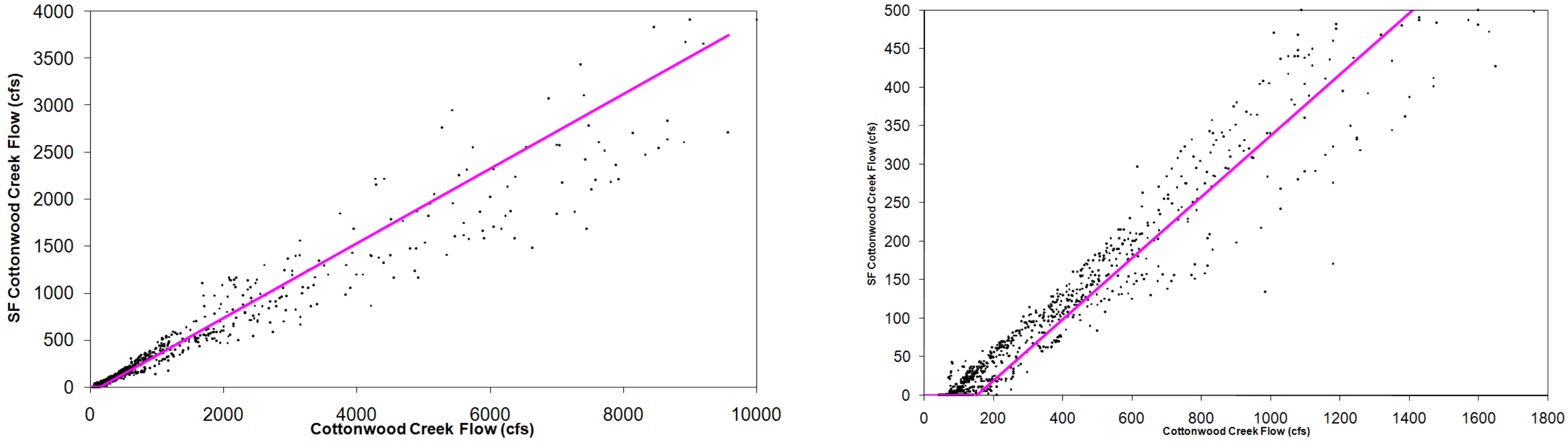


Figure 5
South Fork Cottonwood Creek Flow Data and Regression

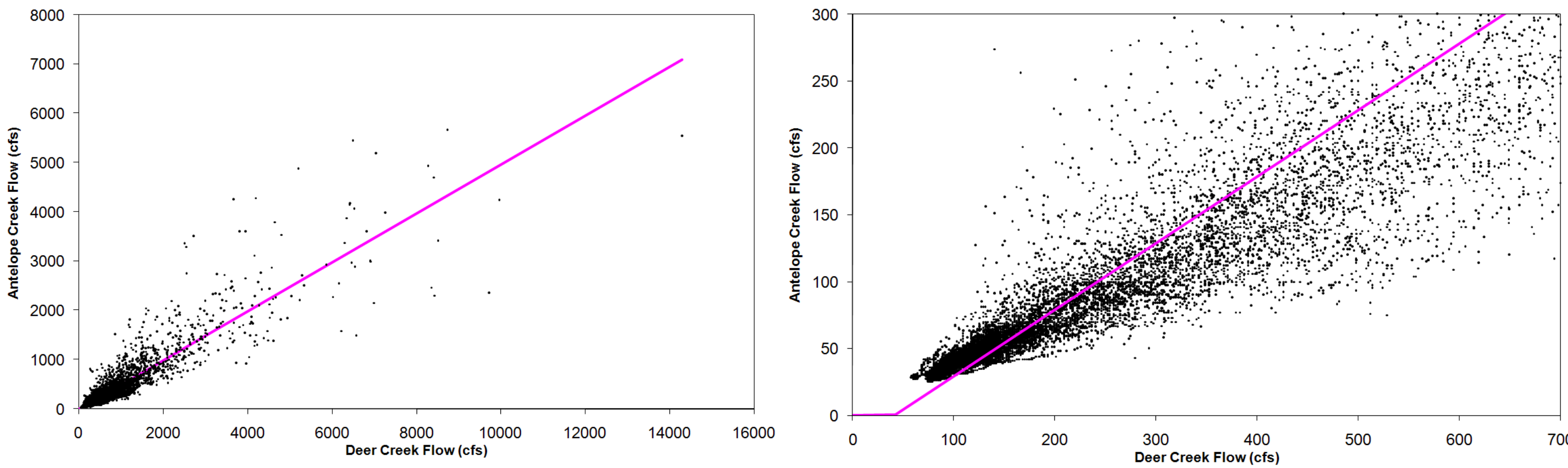


Figure 6
Antelope Creek Flow Data and Regression

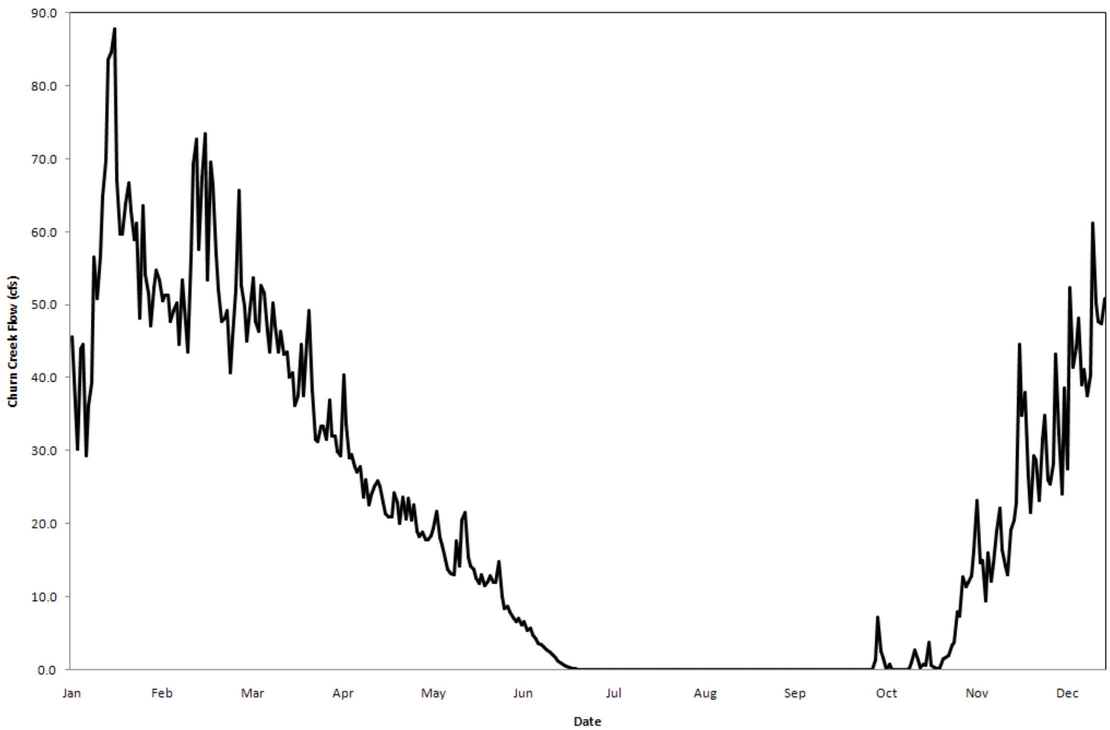


Figure 7
Churn Creek Average Annual Hydrograph

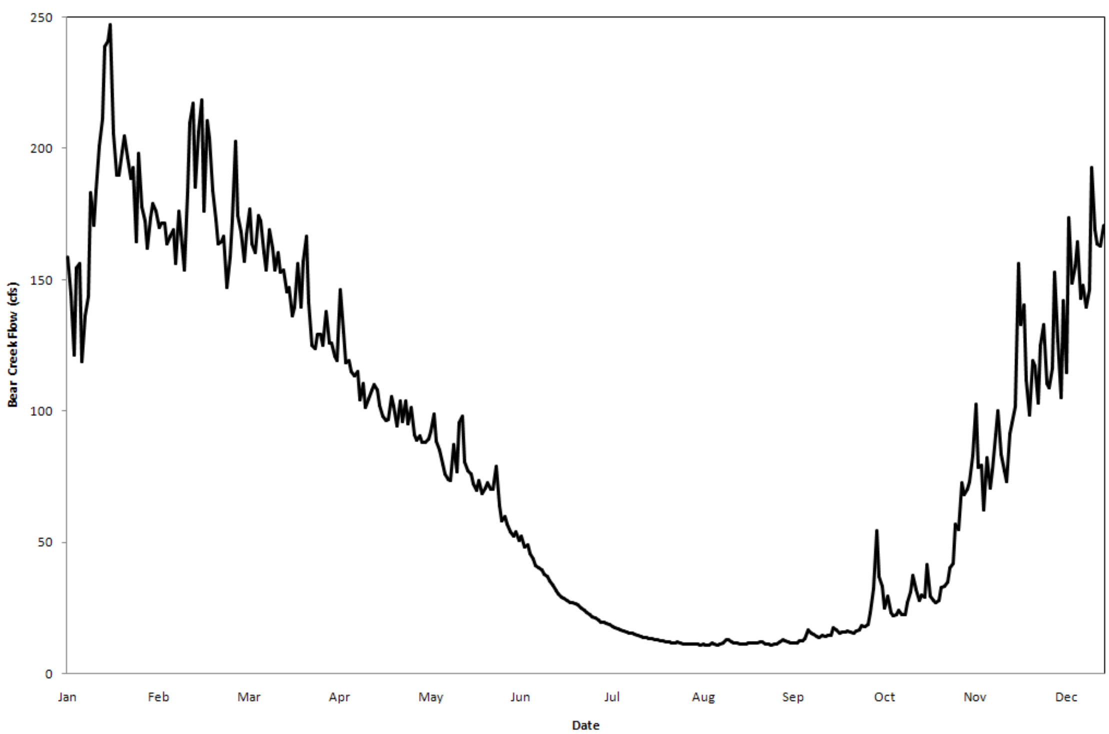


Figure 8
Bear Creek Average Annual Hydrograph

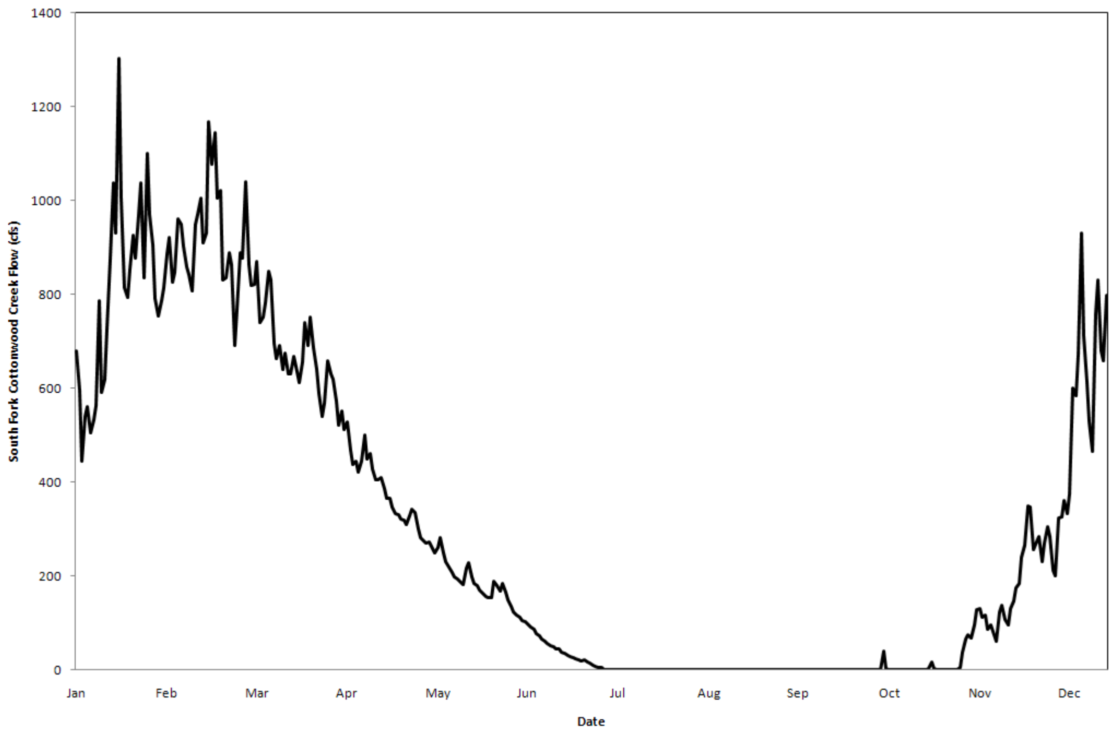


Figure 9
South Fork Cottonwood Creek Average Annual Hydrograph

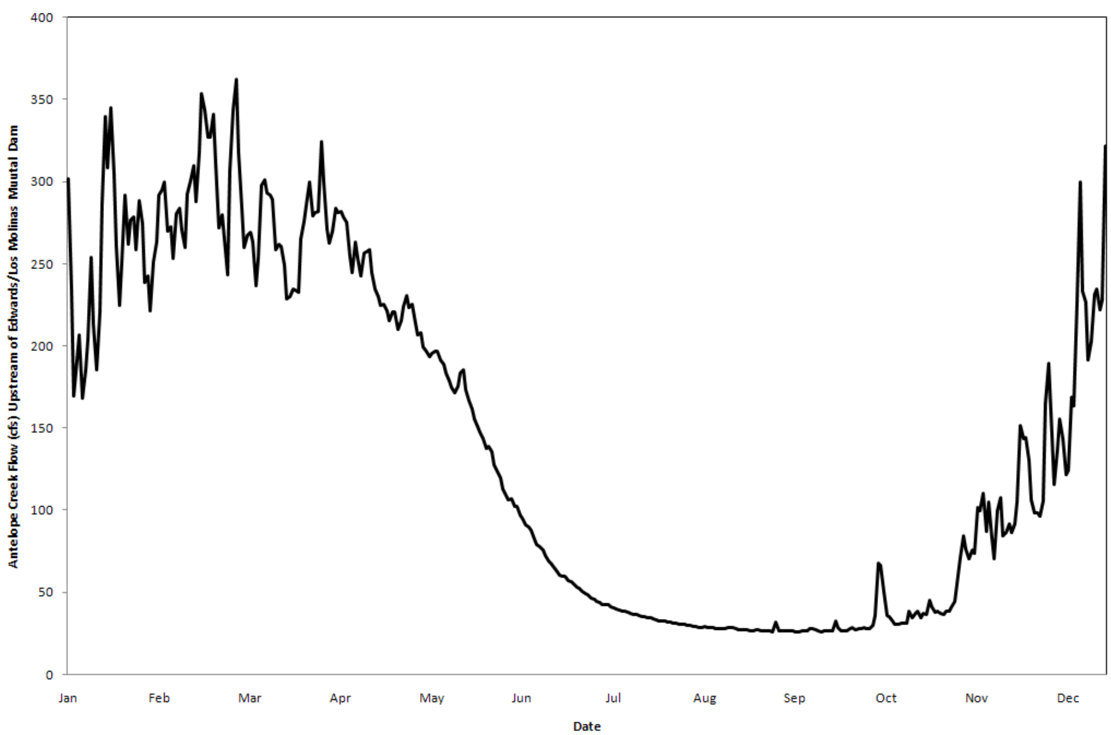


Figure 10
Antelope Creek Average Annual Hydrograph Upstream of Edwards/Los Molinas Mutual Dam

Table 10 summarizes the flow measurements that we made in FY 2009, while Table 11 presents the regression formulas based on the data in Table 10 and gage flows for Cow and Deer Creeks. Figures 11 and 12 show the annual average hydrographs for Stillwater and Antelope Creeks, computed from the period of record flows for Cow and Deer Creeks and the flow/flow regression equations in Table 11. Figures 13 to 15 show the measured flows for Churn, Bear and South Fork Cottonwood Creeks relative to the regression equations computed from historical gage data. Table 12 presents the flow observations we made for New Creek, Craig Creek and Butler Slough.

Figures 16 to 20 show the results of the FY 2009 water temperature monitoring, while Table 13 shows the regression equations we developed from the FY 2009 water temperature data. Web sites for the flow and air temperature data to plug into the regression equations in Tables 9, 11 and 13 are given in Table 14.

Table 10
Fiscal Year 2009 Flow Measurement Data (cfs)

Date/ Location	**Churn Creek**	**Bear Creek**	**Stillwater Creek**	**South Fork Cottonwood Creek**	**Antelope Creek at Highway 99**	**Antelope Creek at upstream end of Cone Grove Park**
Easting[9]	0553312	0575441	0562804	0556046	0575817	0573812
Northing	4499204	4486979	4481286	4468219	4440215	4447042
3/16/09		79.2		184.3	6.3	
3/19/09	28.8		212.8			
5/26/09	1.7	13.8	8.7	58.1	14.4	
6/24/09		10.23		24.6		6
6/25/09	1.56		2.22			
7/23/09	0.16	3.7	0	0.12		1.36
8/26/09		3.95	0			0.287
9/15/09	< 0.1	6.13	0	0		1.23

Table 11
Flow/Flow Regresssions

Regression Equation	**R^2**
Stillwater Creek Flow = $10^{(-2.05 + 1.61 \times \log(\text{Cow Creek Flow}))}$	0.901
Antelope Creek[10] Flow = Max(0, -9.08 + 0.135 x Deer Creek Flow)	0.983

[9] Eastings and Northings are in UTM Zone 10, NAD 83, meters.
[10] This regression is for Antelope Creek at the upstream end of Cone Grove Park, and can be used to predict Antelope Creek flows downstream of the Edwards/Los Molinos Mutual diversion dam.

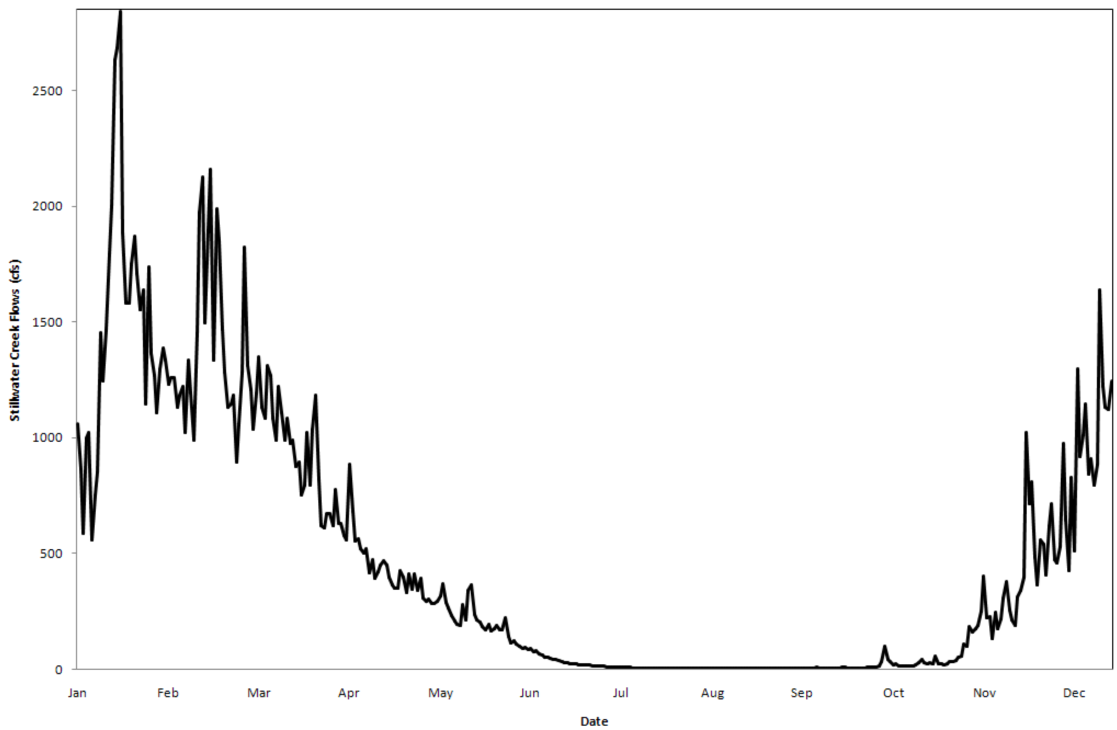


Figure 11
Stillwater Creek Average Annual Hydrograph

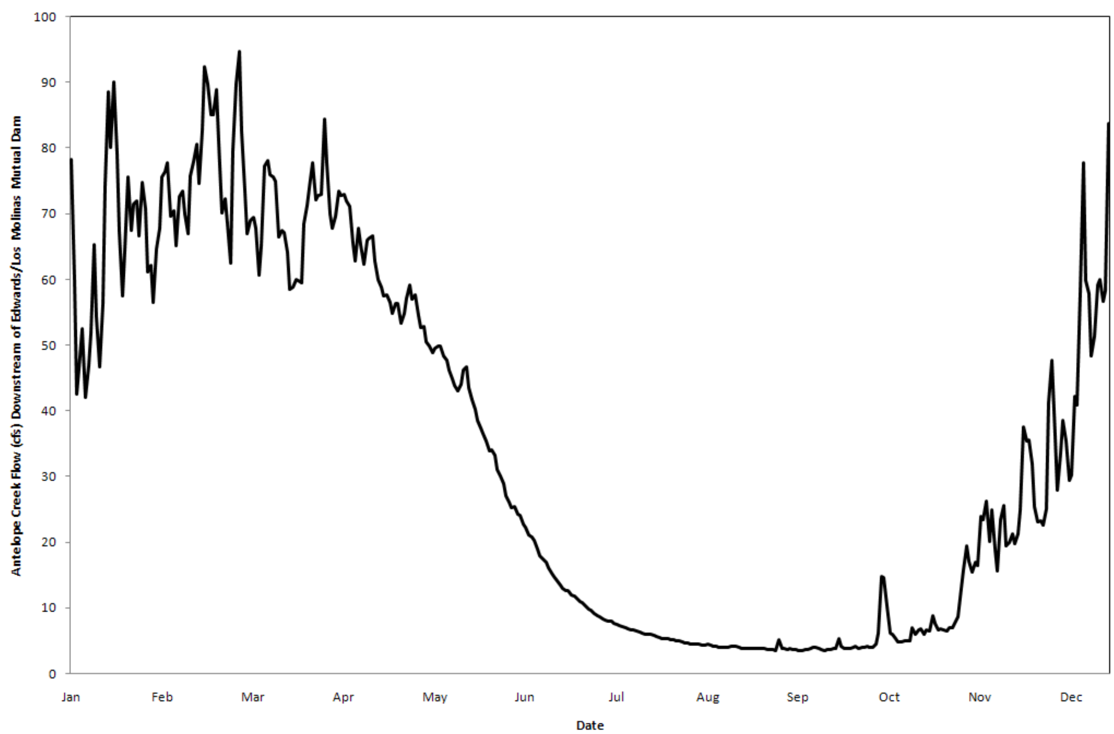


Figure 12
Antelope Creek Average Annual Hydrograph Downstream of Los Molinas Mutual Dam

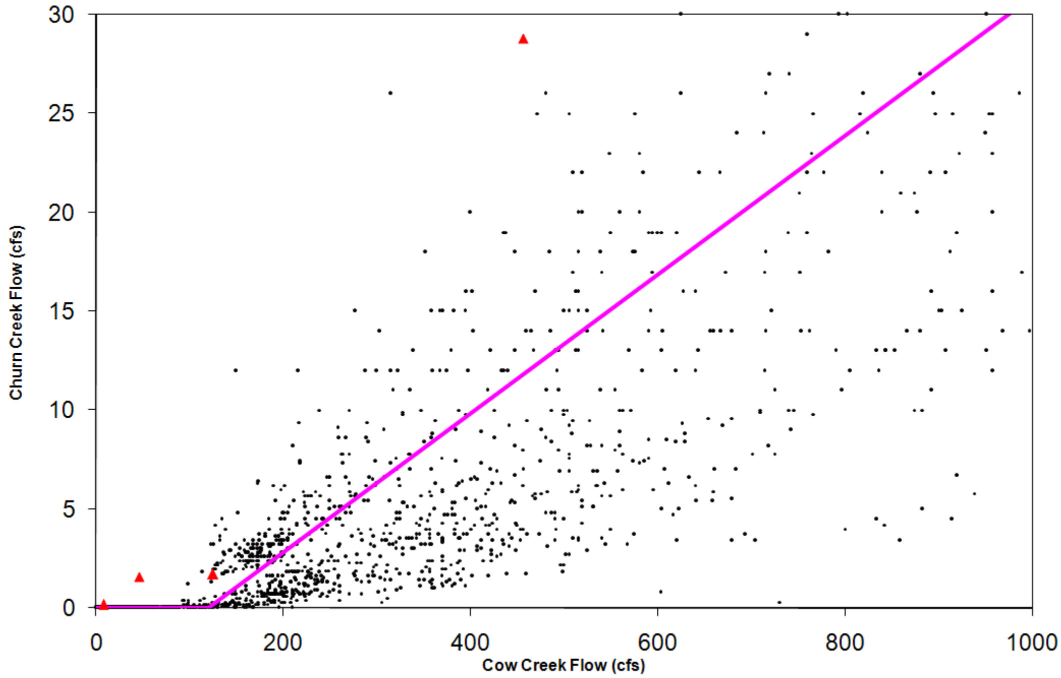


Figure 13

Churn Creek Flows Measured in FY 2009 (Red Triangles) Versus Historic Gage Data (Black Dots) and Flow/Flow Regression Calculated from Historic Gage Data (Purple Line)

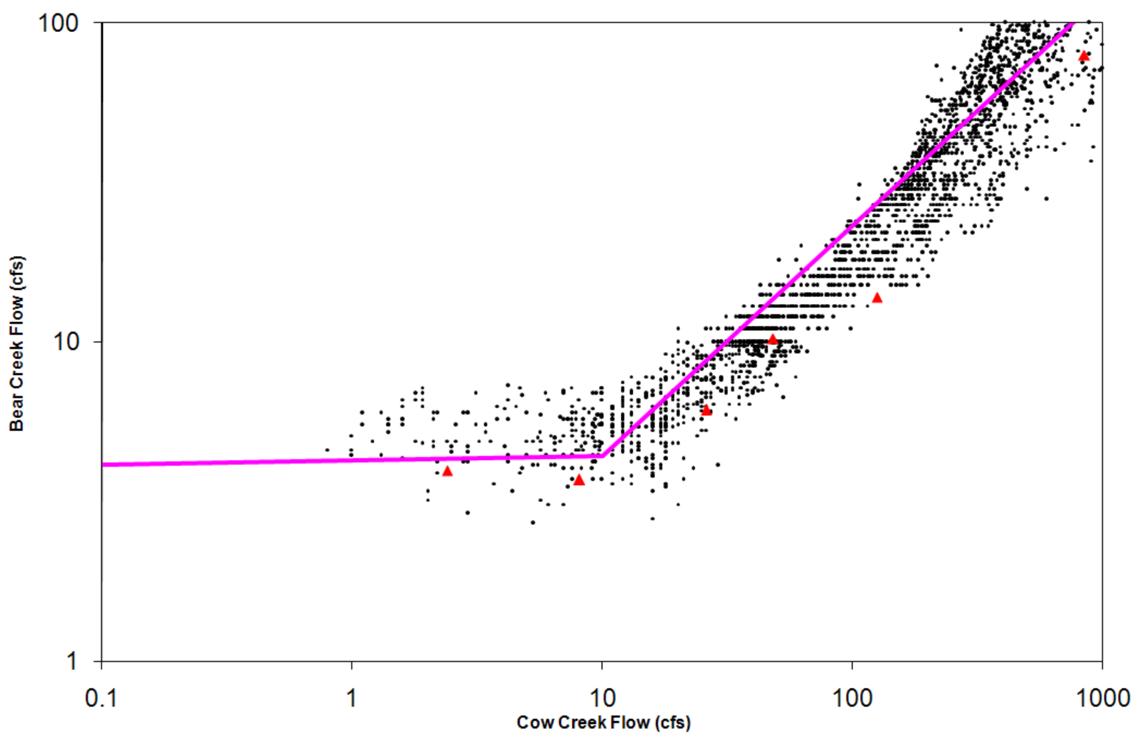


Figure 14

Bear Creek Flows Measured in FY 2009 (Red Triangles) Versus Historic Gage Data (Black Dots) and Flow/Flow Regression Calculated from Historic Gage Data (Purple Line)

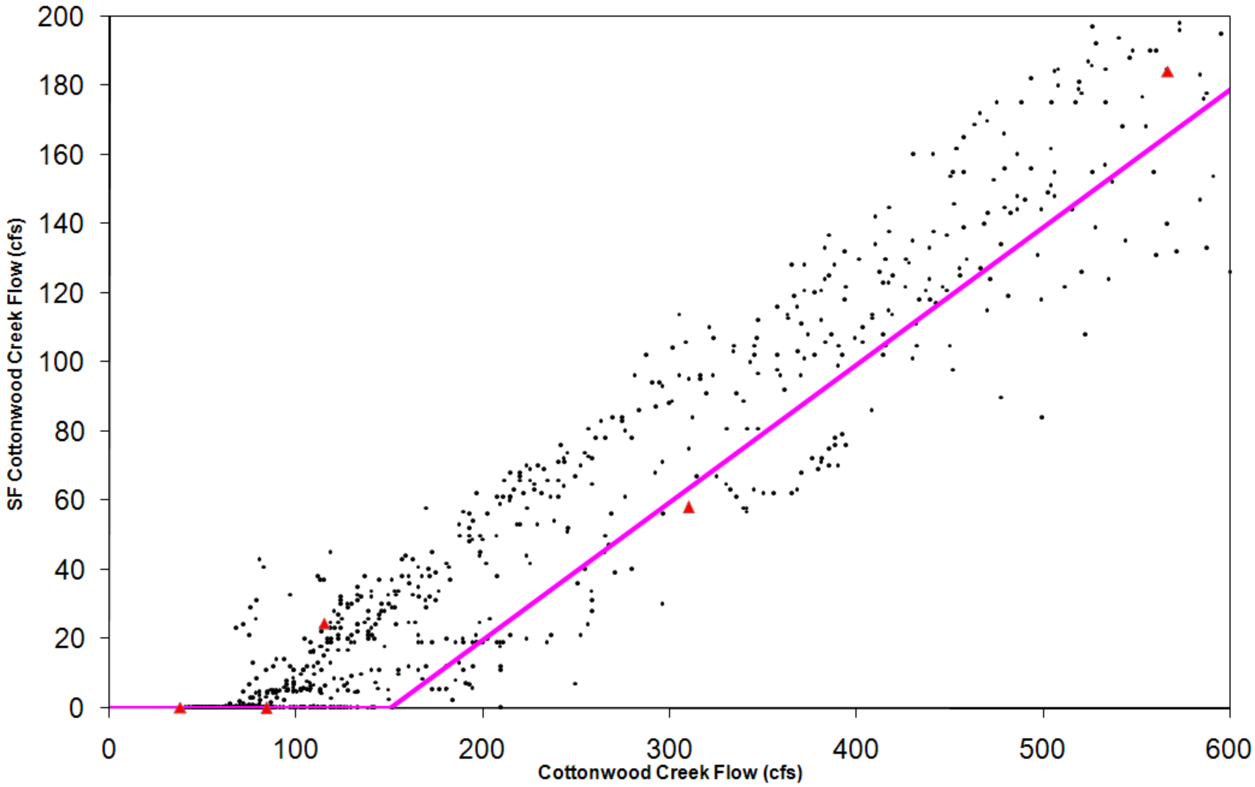


Figure 15
South Fork Cottonwood Creek Flows Measured in FY 2009 (Red Triangles) Versus Historic Gage Data (Black Dots) and Flow/Flow Regression Calculated from Historic Gage Data (Purple Line)

Table 12
Fiscal Year 2009 Flow Observation Data

Date	New Creek	Craig Creek	Butler Slough
5/26/09	Yes	Yes[11]	Yes
6/24/09	Yes	Yes	Yes
7/23/09	Yes	No	Yes
8/26/09	Yes[12]	No	No
9/15/09	Yes	No	Yes

[11] Most of the flow in May and June and all of the flow in July through September was coming from a pipe located approximately 30 feet downstream of the Highway 99 bridge.
[12] All of the flow in August was agricultural return flows.

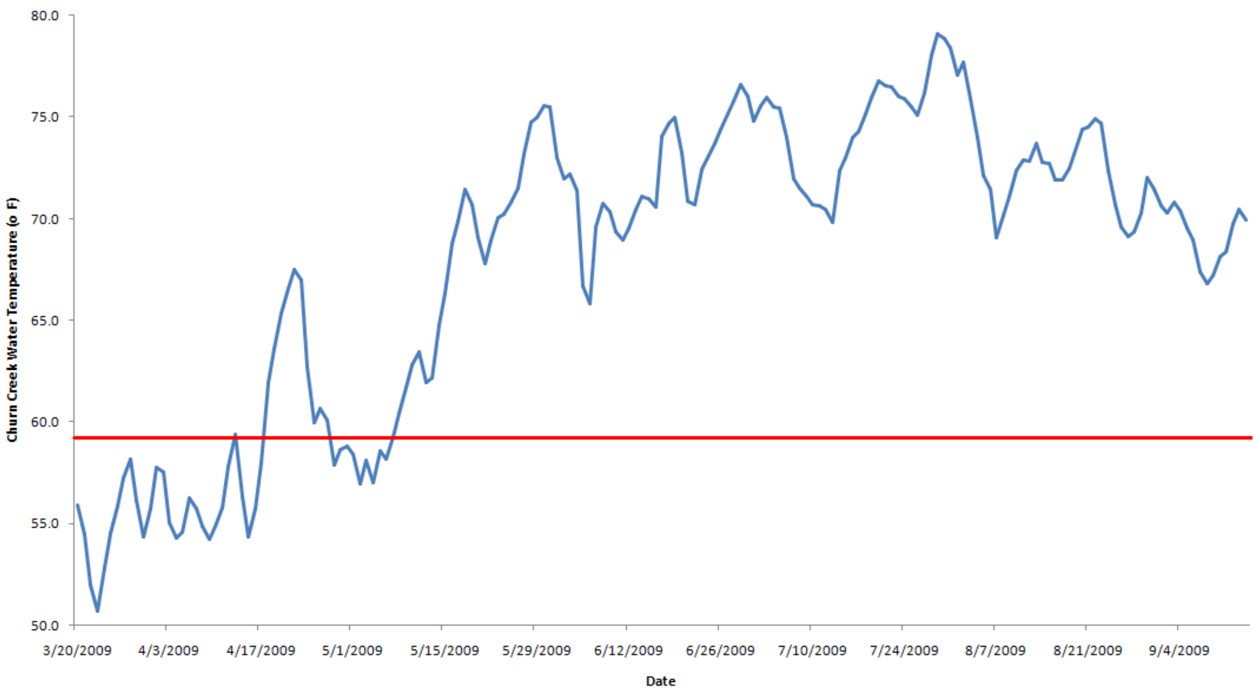


Figure 16
Churn Creek Water Temperatures in FY 2009 Versus 59° F Threshold For Smolt Survival (Mesick 2009)

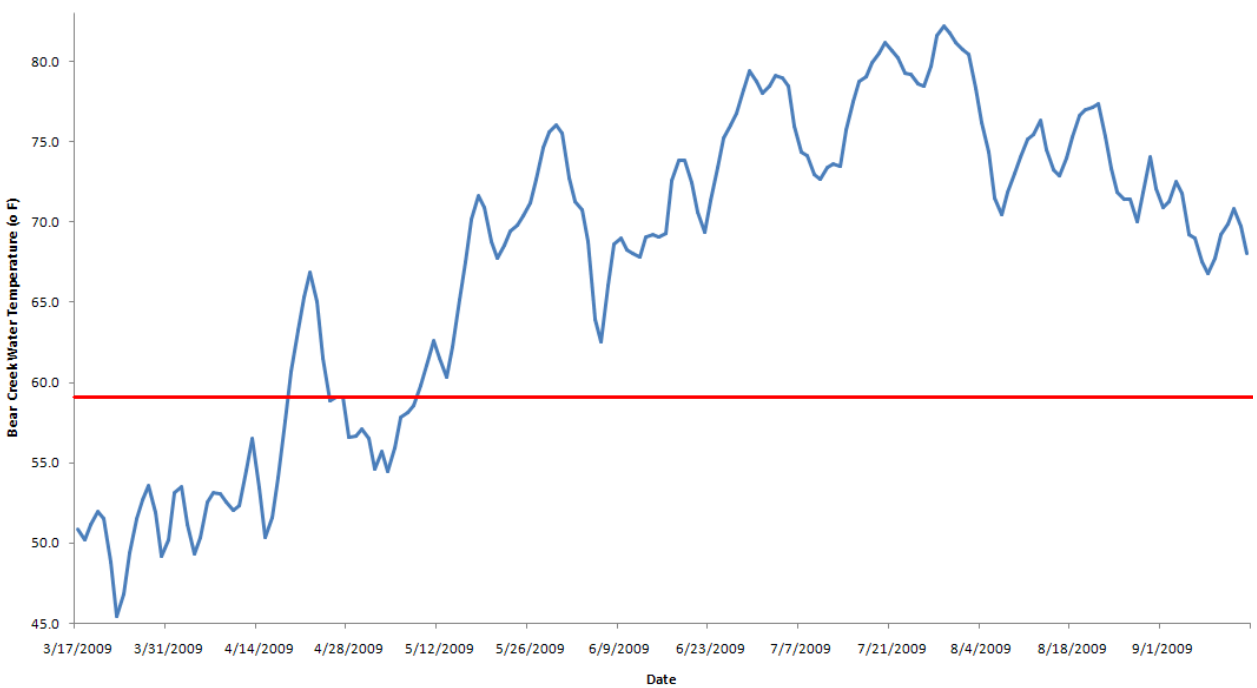


Figure 17
Bear Creek Water Temperatures in FY 2009 Versus 59° F Threshold For Smolt Survival (Mesick 2009)

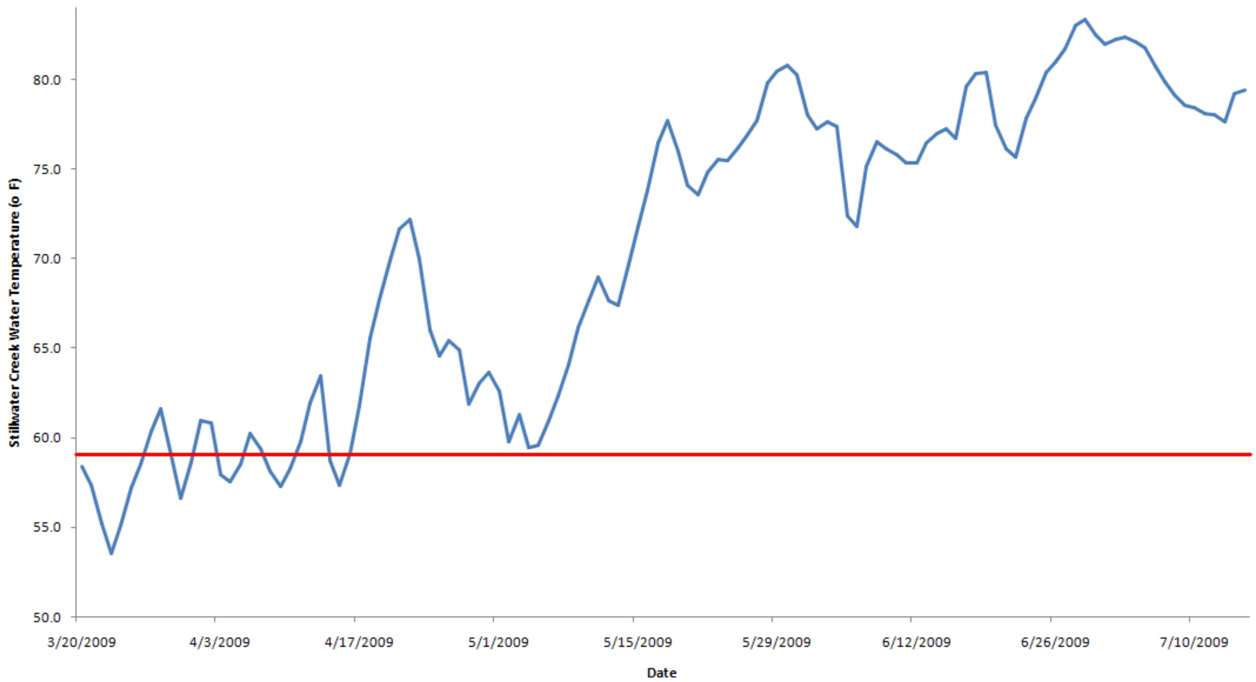


Figure 17
Stillwater Creek Water Temperatures in FY 2009 Versus 59° F Threshold For Smolt Survival (Mesick 2009). Stillwater Creek flows dropped to zero after 7/15/09.

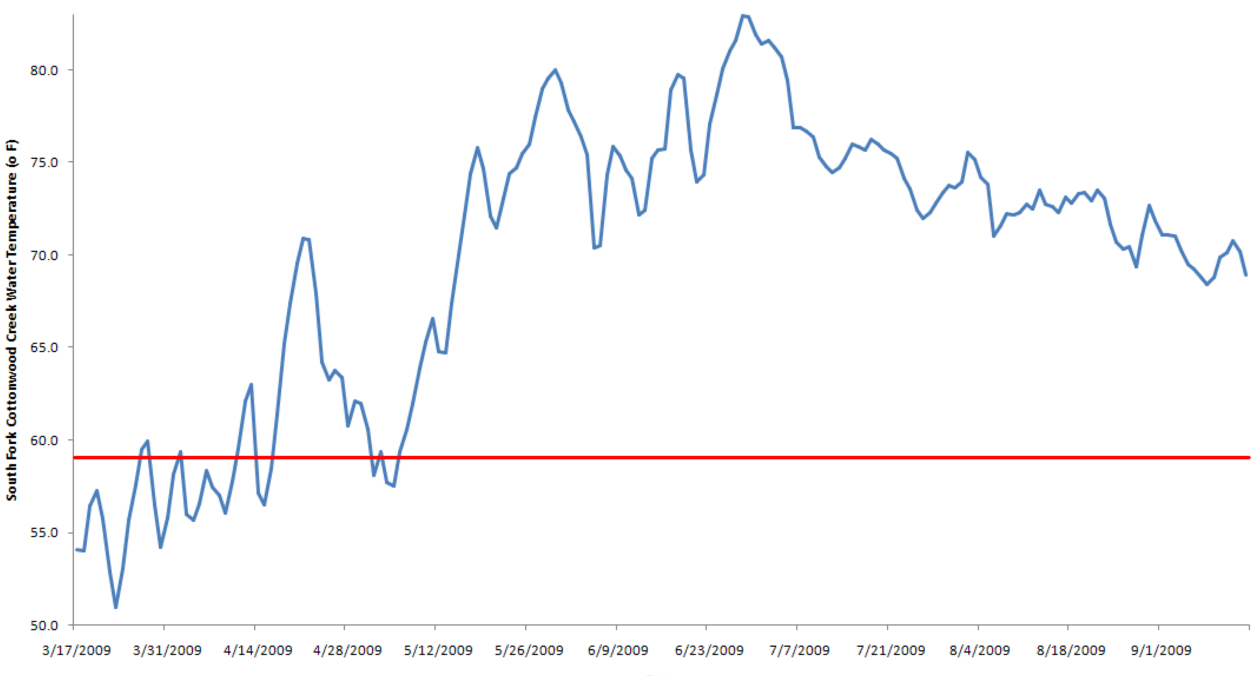


Figure 18
South Fork Cottonwood Creek Water Temperatures in FY 2009 Versus 59° F Threshold For Smolt Survival (Mesick 2009)

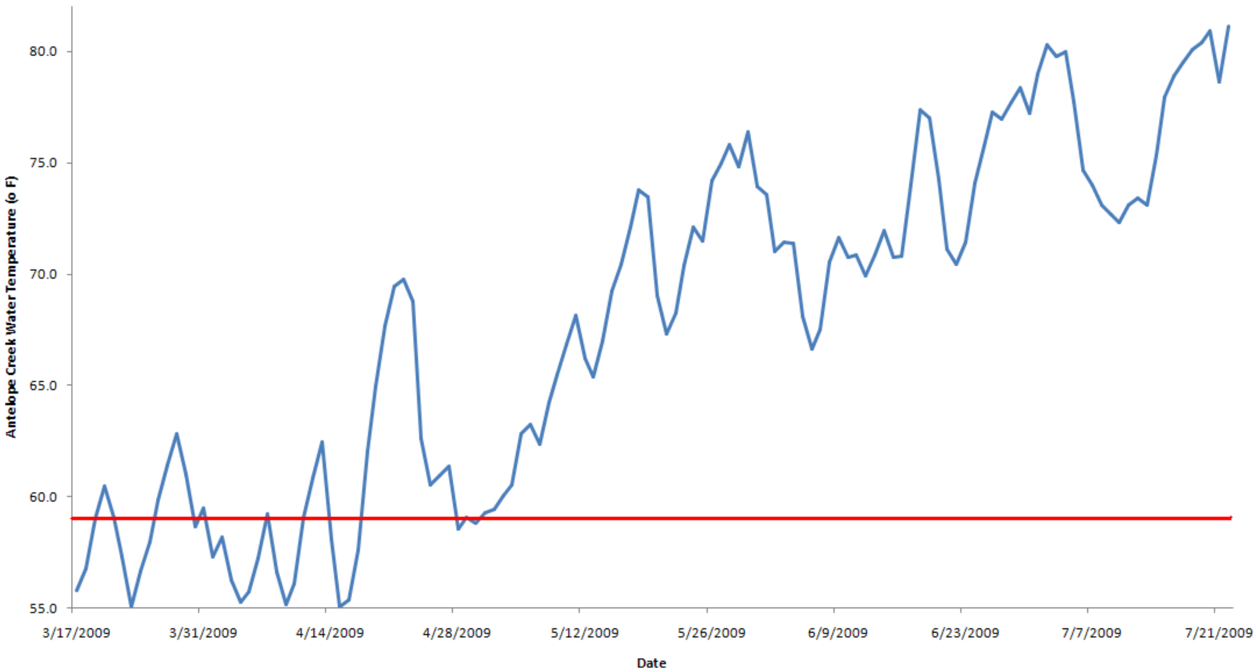


Figure 19
Antelope Creek Water Temperatures in FY 2009 Versus 59° F Threshold For Smolt Survival (Mesick 2009)

Discusssion

Flows for South Fork Cottonwood and Churn Creeks dropped to very low levels starting in July; for this time period, flow in South Fork Cottonwood Creek was irrigation percolation into the streambed while flow in Churn Creek was water percolating from irrigated landscape. As a result, flows during this period give an incorrect picture of what is baseline for the watershed and likely do not provide fish access from these creeks. The regression formula for Stillwater Creek in Table 11 should only be used for Cow Creek flows greater than 26 cfs; based on the data we collected in FY 2009, we would predict that the flow in Stillwater Creek would be zero for Cow Creek flows of 26 cfs or less.

Churn Creek flow measurements in FY 2009 were generally consistent with historic gage records with the exception of the flow measurement on March 19. This flow measurement was likely influenced by an extreme rain event on March 16 that was centered on the Churn Creek watershed. Thus, we do not see this measurement as suggesting that the Churn Creek flow/flow regression no longer applies to Churn Creek. However, we would recommend additional flow measurements to confirm that the Churn Creek flow/flow regression still applies to Churn Creek, since the remaining measurements were taken at very low flows. It is possible that summer flows in Churn Creek have increased since the 1960s due to water percolating from irrigated landscape, since the flow/flow regression predicted that Churn Creek flows would be zero from June through September 2009. If this is the case, the low end of the Churn Creek flow/flow

Table 13
Water Temperature Regresssions

Regression Equation	R^2
Churn Creek Water Temp = 36.5 + 0.451 x Air Temp – 0.012 x Cow Creek Flow	0.866
Bear Creek Water Temp = 24.3 + 0.620 x Air Temp – 0.0137 x Cow Creek Flow	0.887
Stillwater Creek Water Temp = 35.2 + 0.558 x Air Temp – 0.0163 x Cow Creek Flow	0.889
South Fork Cottonwood Creek Water Temp = 34.8 + 0.497 x Air Temp – 0.00439 x Cottonwood Creek Flow	0.742
Antelope Creek Water Temp = 32.3 + 0.534 x Air Temp – 0.0102 x Deer Creek Flow	0.892

Table 14
Web Sites for Data to Plug in to Equations in Tables 9, 11 and 13

Parameter	Web Site
Cow Creek Flows	http://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=html&site_no=11374000&referred_module=sw
Cottonwood Creek Flows	http://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=html&site_no=11376000&referred_module=sw
Deer Creek Flows	http://waterdata.usgs.gov/nwis/dv?cb_00060=on&format=html&site_no=11383500&referred_module=sw
Air Temperatures	http://cdec.water.ca.gov/cgi-progs/queryDaily?s=RED

regression (i.e. Cow Creek flows less than approximately 150 cfs) would no longer result in accurate estimates of Churn Creek flows. Although the flow/flow regression for Bear Creek consistently overestimated Bear Creek flows, relative to measurements taken in FY 2009, the Bear Creek flows measured in FY 2009 fell within the range of historical gage flows. Similarly, the South Fork Cottonwood Creek flows measured in FY 2009 fell within the range of historical gage flows. Thus, it appears likely that the Bear and South Fork Cottonwood Creek flow/flow regressions still apply to Bear and South Fork Cottonwood Creeks; additional flow measurements would help to confirm this conclusion.

Water temperatures in FY 2009 exceeded a 59° F threshold for smolt survival (Mesick 2009) on 4/13, 4/19, 3/27, 3/27 and 3/20 for, respectively, Churn, Bear, Stillwater, South Fork Cottonwood and Antelope Creeks. This suggests that Churn and Bear Creeks may be better choices for restoration activities than the other three Sacramento River tributaries, since water temperatures stay below the 59° F threshold longer. The water temperature regression equations all showed a negative relationship between water temperature and flow, i.e. water temperatures were lower at higher flows. This suggests that water temperatures in these tributaries will stay in an acceptable range for a longer period in wetter years, versus FY 2009. We recommend that additional water temperature data be collected in FY 2010 to verify the water temperature regression equations in Table 13.

Red Bluff Interim Pumping Plant Screens Hydraulic Evaluation

Methods

On June 1 through 11, 2009, an interagency team, with representatives from the Service, National Marine Fisheries Service and the CDFG, measured near-screen velocities on the 10 cone screens located on the intake for the Red Bluff Interim Pumping Plant (Appendix B). Approach and sweeping velocities were measured with a SonTek 16 Mhz Acoustic Doppler Velocimeter (ADV) provided by the CVPIA Anadromous Fish Screen Program. The ADV measured near-screen velocities 3 inches from the screen face. Velocities were measured at 48 locations, in an array of 6 depths and 8 positions around each screen. Velocity measurements were recorded at a rate of 25 HZ for a minimum of 60 seconds.

Results

Approach velocities on screen numbers 6 – 10 had a fairly even distribution of flow through the entire screen area, with no single point exceeding 0.45 ft/s. Flow distribution on screen numbers 1 – 5 were heavily influenced by river current. Approach velocities in areas receiving direct impact of the current far exceeded the design target value of 0.35 ft/s.

Discussion

We recommend that three complete sets of additional velocity measurements be made on the Red Bluff Interim Pumping Plant screens in FY 2010 under a range of different Sacramento River flows and pumping plant operations. The flow and velocity information obtained at the cone screens will help fishery managers assess whether modifications of baffles have ameliorated impediments (e.g. impingement), caused by the operations of the interim pumps, to the downstream migration of various federally listed fish species.

COMPARISON OF PHABSIM AND RIVER2D MODELS

We published a paper (Appendix C) in the January 2009 issue of the International Journal of River Basin Management presenting a comparison of spawning habitat predictions of PHABSIM and River2D from our 1995-2001 CVPIA-funded studies on the Merced, American and Sacramento Rivers. The paper presents the flow-habitat relationships and biological validation results of PHABSIM and River2D.

REFERENCES

Guay, J.C., D. Boisclair, D. Rioux, M. Leclerc, M. Lapointe and P. Legendre. 2000. Development and validation of numerical habitat models for juveniles of Atlantic salmon (*Salmo salar*). Canadian Journal of Fisheries and Aquatic Sciences 57:2065-2075.

Mesick, C.F. 2009. Direct testimony of Carl F. Mesick on behalf of the U.S. Fish and Wildlife Service. Before the U.S. Federal Energy Regulatory Commission Office of Administrative Law Judges. Turlock Irrigation District and Modesto Irrigation District, New Don Pedro Project, FERC Project Nos. 2299-065 and 2299-053. September 11, 2009. U.S. Fish and Wildlife Service: Sacramento, CA.

U.S. Fish and Wildlife Service. 2007. Flow-habitat relationships for spring-run Chinook salmon and steelhead/rainbow trout spawning in Clear Creek between Whiskeytown Dam and Clear Creek Road. U.S. Fish and Wildlife Service: Sacramento, CA.

APPENDIX A

Review of Statistical Approach for Clear Creek Biovalidation

Evaluation Methods for Chinook Salmon Habitat Models – A Literature Review

By

Specializing in Ecological Field Studies

James Griswold

200 S. Second St., Suite B,

Laramie, WY 82070

(307) 634-1756

jgriswold@west-inc.com

For

U.S. Fish and Wildlife Service – Instream Flow Incremental Methodology Program

2800 Cottage Way, Suite W-2605

Sacramento, CA 95825.

October 2009

INTRODUCTION

A critical part of any Chinook salmon habitat model is the evaluation and validation of model performance. Evaluation of a model asks the basic question: “Does the model in its use of the available data answer the researcher’s modeling goals”. Model validation is perhaps the most important stage in the model building process but is often overlooked. Validation is the process used to demonstrate that the model produces reliable output. This paper surveys the most relevant literature on techniques used to evaluate and validate Chinook salmon habitat models. The review identifies the strengths and weaknesses of the evaluation/validation methods used in the literature reviewed, discusses the appropriateness of the various tests for assessing model adequacy and draws conclusions on which techniques are regarded as the most effective. Also reviewed are selected journal articles of non-Chinook salmonid species whose habitat requirements are similar and when evaluation/validation methods are considered to be relevant to those used in Chinook habitat models. Based on the literature and the author’s experience, recommendations are made on identifying an optimal approach for model evaluation and validation.

The paper is organized into the following sections: parametric tests, non-parametric tests, analysis of residuals, ad-hoc methods and simulation techniques. Care will be taken to examine the relevance of the given method to the modeling component being considered by the researcher in the given journal article (ie. whether evaluation is being done at the sampling stage, for the purpose of verifying the hydrodynamic component or biological component, overall goodness of fit, or model comparison). When closely related tests to those found in the literature seem relevant and useful these are reported with key authors being cited. An indexed table (Table 2) summarizing the review by method, journal article and test objective is given to help reference the reviewed material. Variance components related to model evaluation/validation methods will be noted.

Section 1 – Parametric Techniques

Section 1.1 - Pearson correlation Coefficient (r) and R^2

Pearson’s product-moment correlation coefficients (r_p) and R^2 measures are single number descriptors of the degree of linear association between paired samples and model fit respectively. R^2 , in the case of linear models, expresses the fraction of variation in the response explained by the predictors. R^2 used in logistic regression takes a number of forms requiring the user to check software documentation to be sure which R^2 is actually being used. This is essential as the meaning and interpretation of R^2 applied to logistic regression models varies among the different forms. Since r and R^2 are narrowly focused on only a single aspect of the model/data relationship other methods of model evaluation and model validation should also be used.

Ward et al. (2009) used correlation and R^2 to test for a relationship between invertebrate biomass and canopy shading and to determine whether the variation in prey biomass was confounded with salmon stocking density or loss. In the Ward study confidence intervals are reported for R^2. Palm et al (2009) used R^2 in a linear regression of minimum winter habitat suitability for brown trout and the percentage (arcsine square root transformed) of tagged trout remaining in their site of origin. The correlation coefficient and R^2 are applied in instream flow incremental (IFIM) models (Bovee 1978, Bovee 1998,Wood 2009) to assess and validate the calibration of linear regression models in which data input as discharge, temperature, depth, velocity, substrate and cover is predicted using gauging station data or other sampled data. Some background on the concept of IFIM will be given to clarify the context in which r and R^2 are used in this type of model. The Instream Flow Incremental Method (IFIM) was developed by personnel of the Cooperative Instream Flow Service Group, U.S. Fish and Wildlife Service, Fort Collins, Colorado. IFIM allows quantification of the amount of potential habitat available for a species and life history phase, in a given reach of stream at different channel configurations, slopes, water velocities, depths, substrates and other physical variables (Bovee 1978). IFIM is composed of a library of linked models that characterize the spatial and temporal features of habitat resulting from a given river regulation alternative where the model is adaptive and can be tailored to specific needs (Bovee et al. 1998).

One source of input data for hydrodynamic models is discharge. The amount and type of error in discharge measurement in space and time is certain to affect the output of the final model. Gauging station error analysis is discussed in the course book on IFIM developed by Bovee (1996). In many cases the hydrologic input component of the model is to be evaluated under situations in which semi-permanent stream gauges are installed and calibrated later to be used to relate river stage to discharge within the study reach. Evaluation of model input data at this stage consists of developing and testing various regression models relating discharges between the semi-permanent and long-term gages (Bovee 1996). A least squares linear regression is performed between the logarithms of the stage and the logarithms of the discharge. An indicator of the overall quality of a gage station regression model is a goodness of fit criterion such as R^2 or adjusted R^2 (R^2_{adj}) for multiple regression models. In this context R^2 is the portion of the variation in log transformed stage explained by the independent variable(s), log transformed discharge (Neter et al. 1996, Mendenhall and Sincich 2004). R^2_{adj} provides an adjustment for the number of independent variables in the model and provides some protection against the effects of model over fitting. The correlation coefficient (r) or multiple correlation coefficient may also be used, where Bovee (1996) recommends that r equal 0.90 or greater where the r is highly significant (e.g., $p < 0.05$). If this criterion for r is not met the author advises using another method or reducing measurement error, though the method by which this could be achieved is not given. It could be argued that this cutoff level for r should also be based on the researcher's expected model performance and the objectives of the study since lower tolerances for final model predictions would necessitate higher values of R^2 and r for acceptable calibration of physical model components. Sample size will also affect the precision of regression models, however the author does not address the key relationship between sample size, model evaluation, and the statistical power necessary to correctly reject a false null hypothesis of non-zero regression coefficients. In other words r and R^2 need to be supplemented with other tests of model adequacy. R^2 and r are examples of model evaluation which fall into the category of re-substitution methods where the data used to fit the model is also used to test it. Re-substitution methods tend to suggest an overly optimistic accounting of model goodness of fit and validation since they optimize only over the error structure of the data on which the model was fit (Neter et al. 1996, Mendenhall and Sincich 2004).

Williams (2009b) suggests the use of logistic regression as an alternative to habitat suitability models. Knapp and Preisler (1999) in developing habitat models to predict Chinook salmon redds applied a logistic regression analysis in which R^2 was reported as a measure of model fit and for the purpose of model comparison. Various measures called R^2 have been proposed for use in logistic regression as a measure of model goodness of fit (Hosmer and Lemeshow 2000, Mittlbock and Schemper 1996, Menard 2000, Shtatland et al. 2000 SAS, Paper 256-25). Analogous to R^2 in linear regression defined as the ratio of explained sum of squares to total sum of squares, R^2 in logistic regression is a measure of proportional reduction in error measure. These measures apply comparisons of the predicted values from the fitted model to those from model (0), a no data or intercept only model. These measures may best be used to compare models fit to the same data, Hosmer and Lemeshow 2000. However McCullagh and Nelder (1989) warn against sole reliance on the Deviance and Pearson's statistic so that use of a measure of R^2 for logistic regression should be considered, Shtatland et al. 2000. According to Mittlbock and Schemper 1996 R2 measures for logistic regression should have three properties: (1) the measure should have an easily understood interpretation (2) the measure can attain a lower bound of 0 and an upper bound of 1 and (3) the measure is consistent with the character of logistic regression (i.e., not being changed by a linear transformation of model covariates). These authors recommend two for regular use: the squared Pearson correlation coefficient of observed outcome with the predicted probability and a linear regression-like sum-of-squares R^2. For a situation with n covariate patterns the squared Pearson correlation coefficient is

$$r^2 = \frac{\left[\sum_{i=1}^{n}(y_i - \bar{y})(\hat{\pi}_i - \bar{\pi})\right]^2}{\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right] \times \left[\sum_{i=1}^{n}(\hat{\pi}_i - \bar{\pi})^2\right]} \tag{1.1}$$

Where $\bar{y} = \bar{\pi} = n_1 / n$. A linear regression-type measure is

$$R_0^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{\pi}_i \right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2} \tag{1.2}$$

Hosmer and Lemeshow 2000 provide versions of the two measures which apply to the case when $J < n$ covariate patterns

$$r_c^2 = \frac{\left[\sum_{j=1}^{J} \left(y_j - m_j \bar{y} \right) \left(m_j \hat{\pi}_j - m_j \bar{\pi} \right) \right]^2}{\left[\sum_{j-1}^{J} \left(y_j - m_j \bar{y} \right)^2 \right] \times \left[\sum_{j=1}^{J} \left(m_j \hat{\pi}_j - m_j \bar{\pi} \right)^2 \right]} \tag{1.3}$$

and

$$R_{0c}^2 = 1 - \frac{\sum_{j=1}^{J} \left(y_j - m_j \hat{\pi}_j \right)^2}{\sum_{j=1}^{J} \left(y_j - m_j \bar{y} \right)^2} \tag{1.4}$$

The R^2 available in SAS: R_{SAS}^2 is defined as follows

$$R_{SAS}^2 = 1 - \exp\left\{ 2 \left[\log L(M) - \log L(0) \right] / n \right\} \tag{1.5}$$

where $\log L(M)$ and $\log L(0)$ are the maximized log likelihood for the fitted model and the "null" model having the intercept only, and n is the sample size, Shtatland et al. 2000. Since this measure cannot attain a value of 1, Nagelkerke (1991) gave the following adjustment:

$$Adj - R_{SAS}^2 = R_{SAS}^2 / \left[1 - \exp\left(2 \log L(0) / n \right) \right] \tag{1.6}$$

which is labeled in SAS as "Max-rescaled RSquare". $Adj - R_{SAS}^2$ has been criticized by Mittlbock and Schemper (1996) in that there is no reason why the scaling for intermediate values of the measure should be adequate. Thus in applications the value of R_{SAS}^2 may be too small and the value of $Adj - R_{SAS}^2$ may be too large. To correct for this shortcoming Shtatland et al. 2000 propose a deviance R^2 as follows

$$R_{DEV}^2 = \left[\log L(M) - \log L(0) \right] / \left[\log L(S) - \log L(0) \right] \tag{1.7}$$

where $\log L(M), \log L(0)$, and $\log L(S)$ are the maximized log likelihoods for the currently fitted, "null", and saturated models correspondingly (Hosmer and Lemeshow (1989), Agresti (1990, Menard (1995)). The essence of R^2_{DEV} is that it compares the log-likelihood gain achieved by the fitted model (the numerator in 1.9) with the maximum potential log-likelihood gain (the denominator in 1.9)), Shtatland et al. 2000. The authors state that the measure can be interpreted in terms of proportionate reduction in recoverable information and since it is a measure of two log-likelihood gains, R^2_{DEV} can be treated as an indicator of goodness-of –fit. Another R^2 analog that is based on the log-likelihood

$$l(y,\pi) = \sum_{i=1}^{n} \left\{ y_i \log(\pi_i) + (1-y_i)\log(1-\pi_i) \right\}$$

where z is the observed value. R_l^2 is given by

$$R_l^2 = 1 - \frac{-l(y,\hat{\pi})}{-l(y,\pi^0)} \tag{1.8}$$

Here the negative log-likelihood is used, indicating that a smaller $-l(z,\hat{\pi})$ corresponds to a better fit. R_l^2 is called pseudo R^2 in Stata's logistic command. R_l^2 = 0 when model (1.3) is the base model and R_l^2 = 1 when $\hat{u}_i = y_i$. Since it is based on the log-likelihood $l(y,\hat{\pi})$,Laio and McGee 2003 point out that the value of R_l^2 cannot decrease when additional predictors are added to the model; this is not true for R_0^2. Based on their experience with extensive simulated data, Laio and McGee 2003 found that r^2 (1.4) and R_0^2 (1.5) are almost identical numerically. Laio and McGee 2003 developed improved $R^2_{l,adj}$ and $R^2_{0,adj}$, Mittlbock and Schemper (1996), with the following

$$R^2_{l,adj^*} = 1 - \frac{-l(z,\hat{\pi}) + (1+p+\varepsilon_p)/2}{-l(z,\hat{\pi}_0) + (1+\varepsilon_0)/2} \tag{1.9}$$

$$R^2_{l,adj} = 1 - \frac{IPE_l^p}{IPE_l^0} \tag{1.10}$$

$$R^2_{0,adj} = 1 - \frac{IPE_o^p}{IPE_o^0} \tag{1.11}$$

where the bias corrected inherent prediction errors are given by

$$IPE_l^p = -n^{-1} l(z, \hat{\pi}) - B_l(\pi) \text{ and}$$

$$IPE_o^p = n^{-1} \sum_{i=1}^{n} (y_i - \hat{\pi}_i)^2 - B_o(\hat{\pi}),$$

Respectively. The bias terms above use independent replications of y_i in logistic regression (1.3). The $R^2_{l,adj}$ and $R^2_{0,adj}$ can be implemented in R, R Core Development Team 2005, using the function R2.adj available from the authors at http://www.geocities.com/jg_liao/software. The $B_l(\hat{\pi})$ and $B_o(\hat{\pi})$ are evaluated using the Monte Carlo method. The Monte Carlo sample size N is chosen so that, with 95% confidence, the desired expectation is estimated with a relative error less than 5%. In simulation studies Liao and McGee 2003 found that R_0^2 and R_l^2 increased drastically when irrelevant predictors were added to the model. $R^2_{l,adj}$ and $R^2_{0,adj}$ were most robust with respect to irrelevant predictors and were closest to the true coefficient of determination . It would seem appropriate that investigators report which specific measures of R^2 are being used as results can vary among measures for the same study. Hosmer and Lemeshow (2000) state that even though logistic R^2 may be low, the model may fit the data well. They caution that people are used to seeing high R^2 in multiple linear regression studies . The authors advise that researchers using logistic regression should use and report other measures of goodness of fit as well as R^2.

Caution should be used in reporting p-values for the Pearson's correlation coefficient based on the *t*-distribution as the assumption that the samples follow independent normal distributions is needed for the *p*-values to be valid (Efron 1982). An alternative is to use Kendall's rho or Spearman's tau to estimate a rank based measure of association (Conover 1999). Non-normal theory confidence intervals for all three of these correlation measures can be obtained through bootstrapping (Efron 1982, Efron and Gong 1983, Rizzo 2008, Manly 1997, Davison and Hinkley 1996). Though not found in any of the journal articles reviewed, the coefficient of partial determination and its square root as a measure of correlation, is useful in multiple regression models of filter out the reduction in R^2 conditional on other variables being included in the model. For example, $r^2_{12,3,4}$ measures how much smaller relatively, is the variability in the conditional distributions of Y_1 given Y_2, Y_3, and Y_4, than it is in the conditional distributions of Y_1, given Y_3 and Y_4 only (Neter et al. 1996).

Section 1.2 - Statistical Hypothesis Testing and P-Values

A number of Chinook and steelhead habitat models employ the use of hypothesis testing and p-values(Wood 2009, Palm et al. 2009). Palm et al. (2009) compare habitat suitability range and explanatory value (R^2, linear regression) of the variation in the proportion of the salmonid population that remained stationary and overwintered within different sites from late summer until late winter by comparing *p*-values. However there is a current trend away from the use of hypothesis testing and p-values (Johnson 1997, Burnham and Anderson 2002, Royall 1999). Johnson (1997) advises against the use of P-values and statistical hypothesis testing since these are often confused with or used instead of model evaluations based on measured effect sizes related to a-priori know levels of change in processes which have proven effects on a given response. Instead the use of confidence intervals is recommended since these provide a probability interval of effect size and containment of zero in the interval indicates $P>0.05$ if the interval is at the level 95%. Confidence intervals give an estimate of uncertainty as well. A 95% confidence interval of (-20, 400) informs us that the parameter is not as well estimated as if the interval was (100, 150). The author distinguishes between statistical hypothesis testing and scientific hypothesis testing. Scientific hypothesis testing should replace statistical hypothesis testing since the former postulates a theory which generates predictions. These predictions are treated as scientific hypotheses, and an experiment is conducted to try to negate each hypothesis. If the results of the experiment refute the hypothesis, that result implies that the theory is incorrect and should be altered or thrown out. Johnson points out that most statistical hypotheses are known a-priori to be false. Estimated magnitudes of effects with their standard errors and other measurements of precision, such as the coefficients of variation, should always be reported if P-values are given. Johnson 1997 encourages conducting the same study at the same time but at different sites to obtain comparable results at different spatial scales. Similarly replicated studies at the same sites over time allow evaluation of temporal effects. Evaluating a model ideally would involve the comparison of the results obtained by different investigators. Meta-analysis provides methods for combining information from repeated studies allowing less reliance on significance testing by investigating replicated studies, Lipsey and Wilson (2000). An important aspect of model evaluation should be to determine the relative importance to the contributions of, and interactions between, several processes (Quinn and Dunham 1983) so that for this purpose estimation becomes more important than hypothesis testing.

Section 1.3 – t-tests and ANOVA, ANCOVA, Asymptotic Normal Theory Confidence Intervals and Prediction Intervals

t-tests, linear regression, ANOVA and ANCOVA are used in a wide variety of salmonid habitat models and are represented in many of the journal articles reviewed. In addition to habitat models in which the dependent variable is continuous, examples were found in which the dependent variable was a proportion or percentage. Wood (2009) applied t-tests to compare survival of brown trout fry across years in physical habitat models of temperature effects. ANOVA was used by Ward et al. (2009) to test for differences in invertebrate biomass across sites and by Palm et al. (2009) to evaluate differences in length between groups of fish and habitat suitability index between tagging sites. In this same study analysis of covariance (ANCOVA) was employed to test whether increased stocking density of salmon fry yielded increased population density where log transformed stocking density was the predictor and stream-year combination was the blocking factor. Gallagher and Gard (1999) used ANCOVA to investigate the relationship between mesohabitat WUA and red numbers from 1989 through 1996 in the Merced River and from 1991 through 1995 in the Lower American River. The dependent ANCOVA variable was number of redds, and the categorical variable was year where the covariate was WUA. McCarthy et al. (2009) used MANOVA to evaluate effects of forest cover, stream temperature, season, and fish age on food consumption and growth efficiency of juvenile steelhead. Here the food consumption dependent variable consisted of several functional groups of invertebrates such as soft bodied larvae, aquatic nymphs, winged insects and others. One of assumptions of MANOVA is that the data follow a multivariate normal distribution. This can be a sticking point when considering use of this method for which the dependent variable is multivariable. Johnson and Wichern (2002) provide tests for determining whether data are multivariate normal. Rosenfeld et al. (2008) applied t-tests, linear regression and ANOVA effectively in assessing the effectiveness juvenile coho streamside artificial side channels. Researchers must ensure that basic assumptions of independence, constant error variance and normality be met as part of the model evaluation/validation process when using linear models. Castleberry et al. (1996) recommended that users of PHABSIM should take sampling and measurement problems into account, and warned that 'Estimates of WUA should not be presented without confidence intervals, . . .'

Section 1.4 – Likelihood Ratio Tests

Likelihood ratio tests were used by Knapp and Preisler (1999) to test the significance of each of the independent variables on the probability of redd presence. This test requires nested models, that is models that can be transformed into the simpler one by fixing one or more parameters.

Section 1.5 – AIC, AICc Information Theoretic Methods

Information theoretic methods have gained acceptance in recent years often replacing or supplementing traditional stepwise and best-subsets model selection and variable ranking. AIC is based on the likelihood of the model with a term that penalizes for number of parameters. The procedure involves selecting the best model in a collection of models based on the one with the lowest AIC value. AICc is the bias corrected form recommended for use when sample sizes are small ($n/K < 40$) (Burnham and Anderson 2002). AIC is calculated as follows (Burnham and Anderson 2002):

$$AIC = -2\log\left(L\left(\hat{\theta} \mid y\right)\right) + 2K \tag{1.12}$$

AICc is given as:

$$AICc = -2\log\left(L\left(\hat{\theta} \mid y\right)\right) + 2K + \frac{2K\left(K+1\right)}{n-K-1} \tag{1.13}$$

Knapp and Preisler (1999) in a logistic regression model relating water depth, water velocity, and substrate size to spawning sites of golden trout used AIC to determine the relative importance of significant variables by adding variables to the model in the order of their associated AIC value, such that the independent variable with the largest AIC value was added first and the variable with the smallest AIC was added last. Peterson et al. (2009) developed a habitat model evaluate a stream classification system for estimating fish response to changing streamflow. To identify the best approximating model Peterson et al. (2009) fit all possible combinations of the predictor variables including quadratic terms and two-way interactions and evaluated the relative support for each model using AICc. McHugh and Budy (2004) used AICc with other methods to choose among competing models which evaluated patterns of redd site selection in relation to physical habitat variables (depth, velocity, and gravel size) using logistic regression and which habitat suitability for two populations of spring Chinook salmon in Idaho.

It should be noted that AIC selects the best model in a set. The researcher still must find the best collection of models that fits the available data. That is AIC cannot be considered a substitute for a fisheries habitat modeling approach.

Section 1.6 – Logistic Regression Goodness of Fit Tests

The logistic regression model is given by

$$\begin{aligned} y_i \sim Bernoulli\left(\pi_i\right) \, with \log it\left(\pi_i\right) = b_0 + b_1 x_{i1} + \ldots + b_p x_{ip}, \\ i = 1, \ldots, n. \end{aligned} \tag{1.14}$$

Knapp and Preisler (1999) used χ^2 test to evaluate goodness-of-fit of a logistic regression which predicted redd site location. The goodness-of-fit between the observed and predicted probabilities of red presence was determined using a χ^2 statistic, which was then compared with a χ^2 distribution with the required degrees of freedom where a small p value would indicate that the model does not provide a good fit to the data. Since it is not clear which χ^2 test the authors were actually using some background is given on definitions and procedures of two of the more commonly used logistic regression goodness of fit tests.

Two standard measures of goodness of fit are the sum of squared Pearson residuals

$$\chi^2 = \sum_{j=1}^{J} \frac{\left(y_j - m_j \hat{\pi}_j\right)}{\sqrt{m_j \hat{\pi}_j \left(1 - \hat{\pi}_j\right)}}, \, where \; \hat{y}_j = m_j \hat{\pi} = m_j \frac{e^{g(x)}}{1 + e^{g(x)}}, \tag{1.15}$$

where $g(x)$ is the estimated logit and the deviance

$$D = \sum_{j=1}^{J} d\left(y_j, \hat{\pi}_j\right)^2 \tag{1.16}$$

where the deviance residual is defined as

$$d\left(y_j, \hat{\pi}_j\right) = \pm\left\{2\left[y_j \ln\left(\frac{y_j}{m_j\hat{\pi}_j}\right) + \left(m_j - y_j\right)\ln\left(\frac{m_j - y_j}{m_j\left(1-\hat{\pi}_j\right)}\right)\right]\right\}^{1/2}, \tag{1.17}$$

where the sign, + or -, is the same as the sign of $\left(y_j - m_j\hat{\pi}_j\right)$. D is the likelihood ratio test statistic of a saturated model with J parameters versus the fitted model with p + 1 parameters and is generally chi-square distributed with $J - p - 1$ degrees-of-freedom, Hosmer and Lemeshow 2000. If J is defined as the number of covariate patterns where a covariate pattern is a distinct set of values taken on by the p explanatory variables we may have $J < n$ or $J \approx n$, the latter case often occurring when there are continuous variables. When $J \approx n$, p-values for these two standard methods of evaluating the goodness of fit of a logistic model are incorrect when using the $\chi^2\left(J - p - 1\right)$ distribution. Hosmer and Lemeshow (2000) resolved this problem by developing the Hosmer-Lemeshow test. The Hosmer – Lemeshow test is a widely used measure of logistic regression goodness of fit and is implemented by many software packages such as SAS and STATA. The test groups the data based on the values of the estimated probabilities. There are two ways of grouping: (1) collapse the table based on percentiles of the estimated probabilities and (2) collapse the table based on fixed values of the estimated probability. Using g = 10 groups for y = 1 estimated expected values are obtained by summing over the estimated probabilities over all sites in each group, the being done for y = 0 where the sum is over one minus the estimated probability. The Hosmer – Lemeshow goodness of fit statistic, $\hat{C}$, is obtained by calculating the Pearson chi-square statistic from the g x 2 table of observed and estimated expected frequencies. From Hosmer and Lemeshow (200), $\hat{C}$ is defined as follows:

$$\hat{C} = \sum_{k=1}^{g} \frac{\left(o_k - n_k^{'}\bar{\pi}_k\right)}{n_k^{'}\bar{\pi}_k\left(1-\bar{\pi}_k\right)}, \tag{1.18}$$

where $n_k^{'}$ is the total number of sites in the k^{th} group, c_k is the number of covariate patterns in the k^{th} decile,

$$o_k = \sum_{j=1}^{c_k} y_j$$

is the number of responses among the c_k covariate patterns, and

$$\bar{\hat{\pi}}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$$

is the average estimated probability. Hosmer and Lmeshow (1980) show that when $J = n$ and the fitted logistic regression model is the correct model, the distribution of the statistic $\hat{C}$ is approximated by the chi-square distribution with $g - 2$ degrees-of-freedom, $\chi^2(g-2)$ and claim that it is likely that $\chi^2(g-2)$ approximates the distribution when $J \approx n$.

A goodness of fit test like one of the above should be an essential feature of any habitat suitability model. The tables of individual observed and predicted values along with their χ^2 values can provide valuable information on where the model fits the data well and where it fits poorly. The Hosmer-Lemeshow test is available as an option in the SAS logistic procedure.

Section 2 – Non-Parametric Validation Methods

2.1 – Using New Data to Validate a Model, Cross Validation and Resubstitution

Resubstitution

Resubstitution as the name suggests tests the model's predictability by comparing predictions of the observations with the observed data based on the data that was used to fit the model. This technique is somewhat "self fulfilling" in that since the model optimized over the particular structure, error pattern and outliers of the given data, the probability of getting good predictions is expected to be higher than testing predicted versus observed values based on independent data not used to fit the model. Using new data to find misclassification error rates is ideal. A second best option is cross validation which leaves out a subset of data, refits the model on the remaining data ("training set") and calculates misclassification rates on the left out data ("test" set) (Breiman and Spector 1992).

Using New Data to Validate a Model

Castleberry et al. (1996) advocate that an adaptive management approach to assessing model adequacy in which active manipulation of flows, including temporary imposition of flows which might be harmful be incorporated. This in essence is a way to come up with new data, possibly outside the range of existing data, for model evaluation for boundary conditions.

Williams (2001) asks the question "How well do the PHABSIM models predict the actual values of depth and velocity within the cells." The author suggests a method which is compatible with the Instream Flow Incremental Methodology (Bovee et al. 1998), and is applicable to either one or two-dimensional versions of PHABSIM or to other models that combine hydraulic models with biological models. The method is as follows: select cells in the study reach randomly with enough randomly distributed measurements of depth, velocity, substrate or cover within the chosen cells to estimate both the means and variances, or other measures of central tendency and dispersion, to obtain a given degree of accuracy. Collect the data over a range of discharges since the models may be intended to evaluate habitat over a range of discharges. Compare the estimated means to model predictions. Display measured and predicted values using scatter plots. Differences between measured and predicted values should be summarized in box plots or error dispersion plots as well as by statistical measures. Predictions of the biological models should also be tested and uncertainties in both aspects of the modeling should be reported.

Thomas and Bovee (1993) identify a central question: can HSC developed in one stream (the source stream) be used to determine the quality and quantity of microhabitat in another stream or different reach (the destination stream or reach). Transferability is defined as the condition in which fish should use higher quality microhabitats in greater proportion than they utilize lower quality microhabitats, if the HSC have correctly identified high and low quality. The authors describe a technique to test the transferability of habitat suitability criteria. A requirement is that it must be possible to identify a moderate number (e.g., 30-60) of locations occupied by the target species in the destination stream or reach, but the authors do not explain how this sample size was determined. The first step in testing criteria is obtaining all of the criteria that are to be tested. Next the destination stream is sampled for locations that are either occupied or unoccupied by the target organism. At each sampling site, the following data are collected at locations that were or were not occupied by the target species:

(1) Occupancy (whether location was occupied or unoccupied),

(2) Species and life stage, if occupied,

(3) Activity, if known (usually not known unless observed directly),

(4) Depth at sampling location,

(5) Mean column velocity at location,

(6) Cover type, if used by target organism (often not known unless observed directly),

(7) Substrate at location, if applicable to criteria being tested,

(8) Nose velocity, if applicable to criteria being tested,

(9) Adjacent velocity, if applicable to criteria being tested, and

(10) Distance to cover, if applicable to criteria being tested.

From the habitat suitability criteria, the microhabitat variables for each location can be classified as being optimal, usable, suitable, or unsuitable by using the frequency distribution method described above. Once optimal, usable, suitable, and unsuitable ranges for each variable have been defined, the composite suitability for each location is classified:

(1) For a location to be optimal, all of its microhabitat components (e.g., depth, velocity, and substrate) must be optimal,

(2) A location is considered usable if one or more of its components is classified as usable, but none are classified lower than usable,

(3) A location is considered suitable if one or more of its components is classified as suitable, but non are classified as unsuitable, and

(4) A location is unsuitable if one or more of its components is unsuitable.

If the suitability criteria are transferable to the destination stream or reach two conditions should be met: (a) there should be proportionately more target organisms in microhabitat classified as optimal than microhabitat classified as usable, and (b) there should be proportionately more target organisms in suitable microhabitat than in unsuitable microhabitat. Null and alternative hypotheses are tested using 2 x 2 contingency tables and a one sided chi-square test given as:

$$T = \frac{\sqrt{N}\,(ad - bc)}{\sqrt{(a+b)(c+d)(b+d)}} \qquad (1.19)$$

where N is the total number of measured locations, *a* is the number of occupied optimal locations, b is the number of occupied usable locations, c is the number of unoccupied optimal locations, and d is the number of unoccupied usable locations. Suitable locations are substituted for optimal locations, and unsuitable for usable to test classifications of suitable and unsuitable microhabitat.

	Optimal	Usable	Total
Occupied	a	b	a + b
Unoccupied	c	d	c + d
Total	a +c	b + d	N

For a set of habitat suitability criteria to be considered transferable, both null hypotheses should be rejected at the 0.05 level of significance. Critical values of T are obtained from the normal distribution tables (Conover 1980).

Cross Validation

A model validation method recommended by Bovee (1996) and used in many fisheries habitat suitability models is the jackknife or leave-one-out cross-validation (LOO). Olden et al. 2002 compared re-substitution versus a jackknife approach to model validation of species distribution logistic regression classification predictions and found the re-substitution method gave biased results whereas the jackknife approach gave relatively unbiased estimates of model performance. The estimated rates of model correct classification are shown to be substantially influenced by species prevalence (i.e., the proportion of sites at which a species is present). This can result in poorly performing models being viewed as powerful. McHugh and Budy (2004) used LOO to test presence absence predictions from a Chinook salmon red site logistic regression model. The authors also used resubstitution and compared error misclassification rates for the two methods in which resubstitution classified 76% of the data; cross validation correctly classified 70%. The procedure runs as follows: the researcher leaves out one observation at a time refitting the model using the remaining $n-1$ points. This "training" model is then used to predict the left out point. Repeating the procedure over all *n* observations of the data and summing the *n* squared leave out errors (observed minus leave one out prediction) gives the prediction error sum of squares (PRESS). Better models are those with smaller PRESS values. PRESS values can be calculated without requiring *n* separate regression runs, each time deleting one of the *n* cases. The error from omitting the *i*th case is called the deleted residual and is given by:

$$d_i = \frac{e_i}{1-h_{ii}} \tag{1.20}$$

Where e_i is the ordinary residual for the ith case and h_{ii} is the ith diagonal element in the hat matrix, $h_{ii} = X_i^{'}\left(X^{'}X\right)^{-1}X_i$.

Prediction R^2 (R^2_{pred}) is the difference between the total sum of squares and the prediction sum of squares (PRESS) expressed as a fraction of the total sum of squares in the linear regression case and is useful in comparing models across sites and time periods since it is calculated with data not included in model calculation. Some software packages such as R and Minitab offer PRESS as part of their regression routine.

Related to LOO is a method called cross validation. Cross validation is a data partitioning method that can be used to assess the stability of parameter estimates, the accuracy of a classification algorithm, and the adequacy of a fitted model (Rizzo 2008). A researcher can partition the data into training test sets. The model is estimated using the data in the training set only, and the misclassification rate is estimated by running the classifier on the test set consisting of a subset of the original data. Similarly, the fit of any model can be assessed by holding back a test set from the model estimation, and then using the test set to see how well the model fits the new test data. A second version of cross validation is "*K*-fold" cross validation, which partitions the data into *K* test sets (test points). The data could be divided into any number of K partitions, so that there are *K* test sets. Then the model fitting leaves out one test set in turn, so that the models are fitted *K* times. In using cross validation to perform model validation the prediction error can be estimated without making strong distributional assumptions about the error variable (Breiman and Spector 1992). An example taken from Rizzo (2008) involves model selection from four regression models. These could be hydrodynamic models used to simulate, predict or calibrate discharge, temperature or water quality or microhabitat (Y) from measured data (X):

1. Linear: $Y = \beta_0 + \beta_1 X + \varepsilon.$
2. Quadratic: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon.$
3. Exponential: $\log(Y) = \log(\beta_0) + \beta_1 X + \varepsilon.$
4. Log-Log: $\log(Y) = \beta_0 + \beta_1 \log(X) + \varepsilon.$

Once the model is estimated, we would like to assess the fit. Cross validation can be used to estimate the prediction errors as follows for k-fold (leave out sets of size k) cross validation:

1. Partition the original sample into K subsamples.
2. Of the K subsamples, a single subsample is retained as the validation data for testing the model, and the remaining K – 1 subsamples are used as training data. Fit the models and compute the predicted responses for the test points $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ for example. Compute the prediction error $\varepsilon_i = y_i - \hat{y}_i$.
3. The cross-validation process is then repeated K times (the folds), with each of the K subsamples used exactly once as the validation data.
4. Compute the estimate of mean of the squared prediction errors $\hat{\sigma}^2 = \frac{1}{kn}\sum_{i=1}^{nk} \varepsilon_i^2$.

The model which has the smallest mean squared prediction error is the model which has the best fit for the data. Peterson et al. (2009) estimated the accuracy of a large scale channel classification model for assessing the potential effects of river regulation and water use on stream fish communities by using 10-fold cross validation in which the observations for the "fold" are chosen at random. The data was randomly placed into 10 groups, data from one group were excluded, the model was fit with data from the remaining nine groups, and the percent of each channel unit type was predicted for the excluded group. This procedure was repeated for each group (10 times) and error was estimated as the difference between the predicted and measured channel unit composition. Peterson et al. (2009) obtained two additional model performance measures from the cross validation analysis: bias, estimated as the mean difference, and precision as the square root of the mean of the squared differences across samples.

In stratified K-fold cross validation, the folds are selected so that the mean response value is approximately equal in all the folds. For example, in a model with a dichotomous dependent variable there would be equal numbers of successes and failures in each fold. Kohavi (1995) found in extensive simulations using moderate to large data sets that stratified ten-fold cross validation was superior to the leave one out method and was as good as bootstrap methods.

Breiman and Spector (1992) found in an extensive simulation study that at least 5 folds should be used, 10 is best and that cross validation worked as well as the bootstrap in model selection. A heuristic given by Huberty (2006) recommended for presence/absence models is to use a ratio of training to testing cases of $[1 + (p\text{-}1)^{1/2}]^{-1}$ where p is the number of predictors.

Deleted residuals should be plotted versus fitted and independent values in order to determine where the model fits poorly and to identify outlying Y observations when ordinary residuals would not identify these (Neter et al. 1996).

Section 2.2 – Bootstrap and Permutation Tests

Manly (1997) nicely captures the idea of bootstrapping in the following, "The essence of bootstrapping is the idea that, in the absence of any other knowledge about a population, the distribution of values found in a random sample of size n from the population is the best guide to the distribution in the population. Therefore to approximate what would happen if the population was resampled it is sensible to resample the sample."

Sampling problems for IFIM which use PHABSIM inherent in representing a reach of river with a set of transects are considered using bootstrap confidence intervals (Davidson and Hinkley 1997, Manly 2002, Rizzo 2008) of WUA based on resampling transects within habitat types by Williams 1996. Williams (1996) samples with replacement from 5 transects each from pool, riffle and glide habitats and computes percentile confidence intervals from the bootstrap replicate samples. The bootstrap is used since sample sizes are small and data is non-normal. The replicates vary widely within habitat types and the confidence intervals are extremely wide. The key problem is high variability in measured physical variables within habitat types. Williams recommends increased sample sizes, more attention to sample design and more intensive studies on ecological relationships between Chinook and its habitat. Williams states that besides considering the number of transects to sample, bootstrap confidence intervals of WUA should account for measurement errors at the transects and the variation in the data used in developing the suitability curves. Areas of the stream that might not be sampled due to complex hydraulic should also be taken into account. Location of transects by professional judgment is not recommended since there can be no measure of variability in results. In a follow up paper, Williams (2009) considers the following model evaluation/validation questions for 1-D models using PHABSIM and IFIM: 'how well do transects represent the study sites, how well do the study sites represent the reach and how well does PHABSIM estimate WUA at the transects?'

Williams used the percentile bootstrap to estimate 95% confidence intervals for WUA curves in Cache LaPoudre River 1-D rainbow trout habitat suitability PHABSIM model. To calculate confidence intervals for the set of curves bootstrap replicate samples of the 107 curves were drawn with replacement from the original curves, with stratification by habitat type and "bootstrap replicate" composite WUA curves were calculated from the bootstrap samples. A similar procedure was used for subsets of the curves. The procedure was repeated 2000 times and for each discharge the interval containing the central 1900 bootstrap replicate composite curves was taken as the 95% confidence interval for the estimate of WUA at that discharge. The author was careful to note that the percentile bootstrap is biased for some statistics or datasets having extreme outliers (Davison and Hinkley 1997), but for the data used the means and medians of the bootstrap samples did not differ much so that an adjustment for bias was not necessary. The transects used in this study were chosen deliberately and are not random selections. Williams admits this and makes the assumption that transect placement approximates what would have occurred if the sample transect placements were random. To simulate errors in the WUA estimates, normally distributed random error terms were generated. (Method not given). The error terms had a mean of 0 and standard deviation of 5 or 10% of the corresponding WUA value, so that, WUA estimates were unbiased, and about two-thirds of the resulting WUA values were within 5 or 10% of the assumed 'true' values. The author acknowledges that the modeled errors were 'probably smaller than the errors in actual PHABSIM studies. The effect of changes in sample size (i.e. the number of WUA curves) was investigated by: '(1) bootstrap samples of reduced size were selected from the full set of transects, with appropriate numbers of curves from each habitat type, or (2) subsets of the curves were selected, with appropriate numbers of curves from each habitat type, and bootstrap samples were drawn from the subsets.' The first approach allows confidences intervals to be comparable over the range of discharge values while second approach which is more realistic such that means and confidence intervals vary from subset to subset so that the results are compared graphically at a single discharge. Results Williams (2009) showed that the bootstrap and conventional confidence intervals were nearly identical with the bootstrap intervals slightly higher (possibly the result of the random selection assumption). In the alternate case, when bootstrap sampling was stratified by habitat category the bootstrap confidence intervals were narrower than the conventional intervals. Williams attributes this to the skewed distribution of WUA values in some habitats. In summary Williams (2009) found that confidence intervals around the composite WUA curves are moderately wide (28% of the mean at the peak for the juvenile curve, 18% for the adult curve), and particularly if errors in the WUA curves are considered, the shape of the composite curve and the slope of the curve at a fixed discharge can be very uncertain. The uncertainty or variance increases as sample size decreases, and with the usual number of transects used in PHABSIM models, it was in general large, even if WUA at the transects was estimated without error. Williams recommends that bootstrap confidence intervals be used to estimate the uncertainty in HSC by similar methods as used for WUA. Williams admits that his study does not simulate the uncertainty arising in most PHABSIM studies from grouping transects within study sites, and extrapolating these results to a much longer reach so that his analyses greatly underestimate the statistical uncertainty inherent in most PHABSIM studies. William urges researchers to use random sampling or random proportional sampling to select transects as these will be representative of the infinite population of all transects and that researchers should determine appropriate sample sizes to achieve desired level of model precision before data collection. Manly (2009) is given as a reference to a way to achieve this using results from completed PHABSIM studies.

Vaughn and Ormerod (2005) recommend the following method to evaluate model overfitting using the bootstrap:

1. Estimate accuracy statistic in the training data.
2. Generate a bootstrap of equal size to the training set by sampling training data with replacement.
3. Fit the model in the bootstrap using the same methods as employed to fit it in the original training data; this includes the same variable selection strategy, where applicable
4. Estimate the accuracy statistic within the bootstrap resample. This simulates an accuracy estimate made with the training data
5. Using the same model as in step 4, predict the species distribution in the original training set and estimate the accuracy
statistic. This simulates the use of independent test data
6. Overfitting = (training data estimate in step 4) – (test data estimate in step 5)
7. Repeat steps 2– 6 for 100–200 bootstraps. Average the values calculated in step 6 to provide the overall estimate of overfitting
8. Subtract overfitting estimate from the training data estimate in step 1 to provide an optimism-corrected value.

The non-parametric model-based bootstrap (Davison and Hinkley 1997, Manly 1997, Efron and Gong 1983, Efron 1982) may be used to obtain unbiased estimates of R^2, *mse*, standard errors and confidence intervals of coefficients and prediction intervals when there is confidence that the model is specified correctly (constant error variance model) (Efron1982, Davison and Hinkley 1997) . The bootstrap is also a good choice for regression evaluation when sample sizes are small and it is difficult or impossible to determine whether or not errors are normal (Davison and Hinkley 1997, Manly 1997, Williams 2000). The bootstrap generates random samples from the empirical distribution of the sample. The model-based resampling in linear regression algorithm proceeds as follows (Davison and Hinkley 1997):

For $r = 1, \ldots, n$,
1 For j = 1, . . . , n,
(a) set $x_j^* = x_j$;
(b) randomly sample ε_j^* from $r_1 - \bar{r}, \ldots, r_n - \bar{r}$; then

(c) set $y_j^* = \hat{\beta}_0 + \hat{\beta}_1 x_j + \varepsilon_j^*$.

2 Fit least squares regression to $\left(x_1^*, y_1^*\right), \ldots, \left(x_n^*, y_n^*\right)$, giving estimates $\hat{\beta}_0^*, \hat{\beta}_{1,r}^*, s_r^{*2}$.
where *s* is the regression *mse*.

A different approach should be taken when error variances may not be constant and the data is a sample from some multivariate distribution (X, Y). Here there is no assumption on the random errors ε_j other than independence and we resample cases. The algorithm proceeds as follows (Davison and Hinkley 1997):

For r = 1, . . . , *R*,

1 sample $i_1^*, \ldots, i_n^*$ randomly and with replacement from $\{1, 2, \ldots, n\}$;

2 for j = 1, . . . , *n*, set $x_j^* = x_{i^*}$, $y_j^* = y_{i^*}$ then

3 fit least squares regression to $\left(x_1^*, y_1^*\right), \ldots, \left(x_n^*, y_n^*\right)$ giving estimates $\hat{\beta}_0^*, \hat{\beta}_{1,r}^*, s_r^{*2}$.

Since regression may be used to predict new values of discharge (not included in the available data) for input to the final IFIM model a method is needed for evaluating the precision of these predictions. Bootstrap prediction errors can be used when the above linear regression model assumptions are suspect. Davison and Hinkley advise using the following bootstrap procedure for the constant error variance case when *M* new observations are to be predicted:

For r = 1, . . ., *R*,

1 simulate responses y_r^* according to the model-based resampling algorithm above;

2 obtain least squares estimates $\hat{\beta}_r^* = \left(X^T X\right)^{-1} X^T y_r^*$; then

3 For m = 1, . . . , *M*,

(a) sample $\varepsilon_{+,m}^*$ *from* $r_1 - \bar{r}, \ldots, r_n - \bar{r}$, and

(b) compute prediction error $\delta_{rm}^* = x_+^T \hat{\beta}_r^* - \left(x_+^T \hat{\beta} + \varepsilon_{+,m}^*\right)$

where the quantity to be predicted is $Y_+ = x_+^T \beta + \varepsilon_+$ and the point predictor is $\hat{Y}_+ = x_+^T \hat{\beta}$ where the prediction error is estimated by (b) above. + indicates a new point to be predicted. A $(1-2\alpha)$ prediction interval for Y_+ is estimated by the empirical quantiles of the pooled δ^*'s .
The bootstrap prediction limits are

$$\hat{y}_+ - \delta^*_{((RM+1)(1-\alpha))},\ \hat{y}_+ - \delta^*_{((RM+1)\alpha)} \text{ where } \hat{y}_+ = x_+^T \hat{\beta}. \tag{1.21}$$

Of course parametric asymptotic prediction intervals can be computed in the usual manner when regression assumptions are met (Neter et al. 1996, Mendenhall and Sincich 2004). Most software packages offer this option as a standard feature of the regression function.

A permutation test (randomization test) is a type of statistical significance test in which a reference distribution is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points. From Manly (2009) a two sample test of the difference in randomization test proceeds as follows:

1. The observed absolute mean difference is labeled d_1.
2. 2. It is argued that if the null hypothesis is true (the two samples come from the same distribution), then any one of the observed values $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$ could equally well have occurred in either of the samples. On this basis, a new sample 1 is chosen by randomly selecting m out of the full set of n + m values, with the remaining values providing the new sample 2. The absolute mean difference $d_2 = |\bar{x} - \bar{y}|$ is then calculated from this randomized set of data.
3. Step 2 is repeated a large number of times ($R - 1$) to give a total of R differences $d_1, d_2, \ldots, d_R$.
4. The R differences are put in order from the smallest to largest.
5. If the null hypothesis is true, then d_1 should look like a typical value from the set of R differences, and is equally likely to appear anywhere in the list. On the other hand, if the two original samples come from distributions with different means, then d_1 will tend to be near the top of the list. On this basis, d_1 is said to be significantly large at the 100α% level if it is among the top 100α % of values in the list. If 100α % is small (say 5% or less), then this regarded as evidence against the null hypothesis.

If the labels are exchangeable under the null hypothesis, then the resulting tests yield exact significance levels (Davison and Hinkley 1997). In Olden et al. (2002) Empirical data is used to introduce a randomization approach for assessing whether the performances of the fish habitat models are statistically greater than expectations based on chance predictions. The test requires creating a null distribution of correct classification rates (CCRs) for a given species by randomly permuting the original observations of occurrences among the lake or stream sites, conducting logistic regression analysis using the randomized species occurrence and the original independent variables, and calculating the jackknifed CCR. The procedure was repeated 999 time and the significance level of the predictive model was calculated as the proportion of random CCRs (including the observed CCR) that were as great or greater than the observed CCR.

Permutation tests can be applied to a wide variety of problems including testing for differences between parameters from two or more distributions, (Davison and Hinkley (1997), Edgington and Onghena (2007), Manly (2009)) and so it is surprising not to find it used more frequently in fisheries habitat modeling studies. The randomization test has an advantage over a nonparametric test like the Mann-Whitney U-test because it allows the original data to be used rather than just the ranks of the data (Manly 2009).

Section 2.3 – Rank Tests, Chi-squared, Tests Using a Confusion Matrix, ROC Curves

In a study comparing spawning habitat predictions of PHABSIM and River2D models, Gard (2009), used the Mann Whitney U test to test for each river and in the case of the Sacramento River for each race of Chinook salmon, if there was a significant difference in the composite suitability index (CSI) predicted by PHABSIM for occupied versus unoccupied cells, and if there was a significant difference in the CSI predicted by River2D for occupied versus unoccupied locations. Kolmogorov-Smironov tests were conducted for each site for each set of suitability criteria to test if there was a significant difference between the PHABSIM and River2D flow-habitat relationships. Here the statistic being tested is the median. Knapp and Preisler (1999) conducted the non-parametric rank based Kruskal – Wallis one-way analysis of variance to test for differences in habitat characteristics associated with cells used and not used by spawning golden. The authors used the Kruskal-Wallis test because the data was not normally distributed, had unequal variances, and normality and variance equivalency could not be accomplished using standard transformations. The Mann-Whitney U-test has been applied in a number of model evaluation settings when sample sizes are small and the normality assumption has not been met yet the data distribution is symmetric (Gard 2009, McHugh and Budy 2004). The Mann-Whitney U-test was used by McHugh and Budy (2004) to compare the depth, velocity, and gravel size (D84) values at sites that were used for spawning during 2001 with those for sites that were not used.

Spearman's rho (r_s) is a rank based non-parametric correlation coefficient that assesses how well an arbitrary monotonic function could describe the relationship between two variables , without making any other assumptions about the particular nature of the relationship between the variables (Conover 1980). In Gallagher and Gard (1999) a variation of Spearman's rho called gamma that adjusts for data with many ties was employed to determine if there was a relationship between Chinook salmon spawning density and predicted WUA at the mesohabitat level in the Merced River.

Evaluation Indices Using a Confusion Matrix – Hirzel et al. 2006

Hirzel et al. (2006) review and compare indices based on presence/absence information include Cohen's Kappa, K_{max}, AUC, and adjusted D^2 in the context of evaluating generalized linear models applied to habitat suitability data. These methods index the degree of agreement between prediction and data. The first step is to choose a habitat suitability (HS) threshold which is intended to separate unsuitable areas where the species or redd should be absent, from suitable areas (HS greater than threshold) where it should be present. A *confusion matrix* is created which enumerates how many presence and absence evaluation points occur in the suitable and unsuitable areas, Figure 1. Other methods using this matrix are described in Fielding and Bell (1997). Among the evaluators based on this matrix is the Cohen's Kappa index K (Agresti, 1990) which is computed as follows:

$$K = \frac{N \sum x_{ii} - \sum x_{i.} x_{.i}}{N^2 - \sum x_{i.} x_{.i}} \qquad (1.22)$$

Where x and N are counts of evaluation points as defined in Fig. 1. *K* varies from -1 to 1, high values indicating a good agreement between prediction and data, and 0 corresponds to random agreement. The results obtained from this method depend on the threshold value that the user chooses. Methods which do not depend on a threshold include K_{max} and area under the curve (AUC). K_{max} is the highest Kappa for threshold values from 0 to 1. AUC is found by plotting, for threshold values from 0 to 1, the proportion of true positive x_{11}/x_1 against the proportion of false positives $x_{12}/x_{.2}$. One computes the area under the curve where an AUC of 0 indicates worse-than-random model, 0.5 (random model) and 1 (best model possible).

		Observed		Margin sums
		Presence	Absence	
Predicted	Presence	X_{11}	X_{12}	$X_{1.}$
	Absence	X_{21}	X_{22}	$X_{2.}$
Margin sums		$X_{.1}$	$X_{.2}$	N

Fig. 1 Contingency table of the model predictions against the actual observations. The x_{ij} represent counts of evaluation points, with $N = \Sigma x_{ij.}$ (taken from Hirzel et al. 2006).

Boyce et al. (2002) found a way to relax somewhat the threshold constraint. Their method consists in partitioning the habitat suitability range into *b* classes (or bins), instead of only two. For each class *i*, it calculates two frequencies: (1) P_i , the predicted frequency of evaluation points:

$$P_i = \frac{p_i}{\sum_{j=1}^{b} p_j} \qquad (1.23)$$

Where pi is the count of evaluation points predicted by the model to fall in the habitat suitability class i and $\sum p_j$ is the total number of evaluation points; (2) E_i, the expected frequency of evaluation points, that is, the frequency expected from a random distribution across the study area. This is given by the relative area covered by each class:

$$E_i = \frac{a_i}{\sum_{j=1}^{b} a_j} \tag{1.24}$$

where a_i is the number of grid cells belonging to habitat suitability class I, or area covered by the class I, and $\sum a_i$ is the overall number of cells in the whole study area. For each class I, the predicted-to expected (P/E) ratio F_i is given by

$$F_i = \frac{P_i}{E_i} \tag{1.25}$$

If the habitat model adequately identifies the species suitable areas, a low suitability class should contain fewer evaluation presences than expected by chance, resulting in $F_i < 1$. On the other hand, high suitability classes should have F_i increasingly higher than 1. The plot of *P/E* against the mean habitat suitability of each class provides an easily accessible interpretation tool. Thus a good model should show a monotonically increasing curve, i.e. F_i increasing as suitability increases. Boyce et al. (2002) rate this monotonic increase by the Spearman rank correlation coefficient between F_i and i. This index varies from -1 to 1. Positive values indicate a model for which predictions are consistent with the presences distribution in the evaluation dataset, values close to zero indicate that the model is unlikely to be different from a random model, negative values indicate an incorrect model, which predicts low quality areas where presences are more frequent. Hirzel et al. (2006) modified the "Boyce Index" by Precision of the Spearman rank correlation could be achieved non-parametric confidence intervals through a randomization procedure or bootstrapping (Efron and Gong 1983, Manly 2009).

Section 3 –The Conceptual Model as a Basis for Validation and Evaluation

Section 3.1

An effective approach to model evaluation and validation should depend on the underlying conceptual model.

Ahmadi-Nedushan et al. (2006) describe and compare the following habitat suitability modeling estimation frameworks: multiple regression, logistic regression, logistic regression, generalized linear models, generalized additive models, artificial neural networks, fuzzy rule based modeling and principal components with respect to their strengths and weaknesses. The authors report on several comparative studies and indicate that artificial neural networks show promise.

Castleberry et al. (1996) would require those using PHABSIM to construct model evaluation and validation that considers the following issues: (1) sampling and measurement problems associated with representing a river reach with selected transects and with the hydraulic and substrate data collected at the transects; (2) sampling and measurement problems associated with developing the suitability curves; and (3) problems with assigning biological meaning to weighted usable area (WUA), the statistic estimated by PHABSIM.

Kondolf et al. (2000) point out the potential problems associated with errors related to different spatial scales used in hydrodynamic and biological model components. The conceptual model for PHABSIM assumes that the data obtained from the transects represents half-way upstream or downstream to the next transect. With this conceptual model, the authors point out, validation consists of measuring the depth, velocity, and substrate at random points in the study reach at alternated discharges and comparing these measurements with the values PHABSIM predicted for those points, where validation should include the habitat variables as well as the WUA. Kondolf et al. 2000 state that if the conceptual model for transect data are treated as samples stratified by habitat types rather than as representing specific areas of the channel, validation will depend on the specifics of the sampling design, but the process will remain the same: model predictions of the joint distributions of depth, velocity, and substrate would have to be compared with independent data. If transect sites are chosen randomly, they will give an unbiased estimate of conditions in the study reach, ‘so that models can be validated at the transects and the streamwise spatial sampling errors estimated separately using statistical methods such as bootstrapping’. Kondolf et al. 2000 urge researchers to report estimates of WUA with standard errors or confidence intervals so that all stakeholders are aware of the uncertainty associated with the estimates.

The choice of model tests should be made in the context of how the model will be applied (Fielding and Bell 1997). The authors state that if the objective is to conserve habitats with high opportunity costs the model should accurately predict species presence. If the model is to be used to predict impacts for endangered species false positives may be more critical. Fielding and Bell put forth the following guidelines:

(1) Decide which data are to be used for the estimation of error. Do not rely on an estimate based on resubstitution of the training data. A more robust estimate will be obtained from independent testing data.

(2) If predictions are to be restricted to a homogeneous region consider a data-partitioning technique. If the predictions are to be tested for their generality use a prospective sample selected via temporal or geographical criteria.

(3) If data-partitioning is to be used consider using more than one approach, ideally including k-fold partitioning or jack-knifing. When deciding on a size for the training set use a heuristic such as that suggested by Huberty (2006), but also take into account any cases:variables constraints imposed by classifier.

(4) Understand the nature of any error measures that are used.

(5) If you wish to determine if a classifier predicts better than chance, use a measure such as Kappa or *NMI*. Recall that *NMI* is less affected by prevalence.

(6) ROC plots avoid the problems associated with threshold effects. If error is to be based solely on confusion-matrix-derived measures consider adjusting the threshold. It is desirable to use *a priori* criteria for deciding on a threshold.

(7) If classifiers are to be ranked, comparisons based on ROC plots are likely to be more robust since they are independent of the values in a confusion matrix.

(8) If the aim is to improve within-region accuracy consider using spatial analysis methods that incorporate the almost inevitable spatial autocorrelation.

(9) If the aim is to improve the predictive success with prospective samples, based on a different region, an attempt should be made to remove the spatial structure from the models.

Additional suggestions by Fielding and Bell (1997) include: if appropriate examine the spatial pattern of the errors and consider using, with caution, post-hoc hypotheses to interpret the patterns; consider weighting errors if there are ecological or economic justifications; be cautious of any statement of model accuracy that does not justify the choice of error measure; if after model validation, the aim is to derive a robust classification rule, all of the available data should be used.

Guisan and Thuiller (2005) discuss the importance of matching the resolution or spatial scale at which sampling takes place and that of the resolution that predictions are to be made.

Evaluating the relevance of composite suitability indexes as a probability of use measure is considered in a test procedure described by Williams (2009b). For each model and site, order all cells by CSI, and divide them into ranks by CSI. Then, plot the percentage of all cells in each rank that are used over the mid-point of its range. Thus, with ten ranks, the percentage of cells with CSI that are greater than 0.9 that are used would be plotted over 0.95, the percentage of cells with CSI > 0.8 and ≤ 0.9 that are used would be plotted over 0.85, etc. Confidence intervals for the plots could then be estimated by bootstrapping (Effon 1982, Manly 1997, Davison and Hinkley 199). If the CSI is a kind of resource selection function, then the plot should approximate a straight line. This line should lie on the diagonal if the percentage of occupied cells is scaled by the constant of proportionality between the value returned by the function and the probability that the cell will be used. If instead the CSI is an index of suitability rather than of probability of use, then the percentage of used cells should increase sharply at higher CSI values (Freeman and Moisen 2008), since fish should select the most suitable spawning habitat available. If neither of these conditions obtains, then either the hydraulic model has not performed well, or the utility of the index should be questioned.

Section 4 – Analysis of Residuals and Model Deviations – Graphical and Numerical Methods

Section 4.1 – Histogram, Normal Probability Plots

Many of the journal articles reviewed performed some sort of inspection of model residuals in order to accomplish one or more of the following: determine model fit, identify outliers, check for independence, normality and non-constant errors. Peterson et al. (2009) assessed goodness of fit of the global (all predictors) model by examining residual and normal probability plots. They also looked for potential temporal dependence by inspecting plots of residuals ordered by sample date for each sample site. If there was no trend in the residuals, they assumed that there was no temporal dependence. To check for goodness of fit of logistic regression models and Poisson regression models Peterson et al. (2009) examined residual and normal probability plots. Geist et al. (2000) used refined nearest-neighbor analysis on digitized Chinook redds to determine whether fall Chinook salmon redds were randomly distributed or if they followed a uniform or clustered pattern. The spatial pattern analysis was also used to determine the distance between redds within any given pattern type. 'Refined nearest-neighbor analysis (Boots and Getis 1988) makes use of the cumulative distribution F(d) to characterize the probability that the nearest neighbor to a red is within a given distance d.' Given a random spatial distribution generated by a Poisson process, Geist et al. (2000) give the expected cumulative distribution function as

$$F\left(d\right)=1-e^{-\lambda\pi d^2},\, d\geq 0 \tag{1.26}$$

Where λ is the intensity of the points within the area, estimated by $\lambda = n / A$ for n points in the area A. The empirical cumulative distribution of distances was calculated from the data set for each distance d and compared with the expected value for that distance. They generated a Monte Carlo confidence envelope around the expected value for each distance d. The empirical cumulative distribution determined from the data is compared with the confidence interval for each d: if the proportion of the nearest neighbors less than distance d is outside the confidence envelope, then the hypothesis that the spatial pattern of the data points resulted from a random process is rejected at the 95% confidence level. The direction of the deviation above or below the confidence envelope indicates whether the non-uniform pattern was closer to a clustered or uniform distribution, respectively.

Section 4.2 – Run Charts

Run charts plot residuals or prediction errors versus location or time and are used to identify places in the model where there is lack of fit or lack of independence (Vardeman and Jobe 1999).

Section 4.3 – Standardized Residuals, Prediction Error, Fitted vs Observed, Outliers, Fitted vs. Independent Variables

Standardized regression errors should be plotted against predicted values and against the independent variable to check for unequal error variances, outliers, lack of independence, bias and model misspecification (Neter et al. 1996). Bias in the predicted values may be indicated if the residuals show a consistent pattern above or below the zero line and unequal error variance (dispersion) may show up as a fan shaped pattern in the residual plot against the fitted values (Mendenhall and Sincich 2004). A histogram of errors should indicate that the distribution of errors are approximately normally distributed. Departures from normality in linear models can also be checked by normal probability plots. Here each residual is plotted against its expected value under normality. A plot that is nearly linear is evidence for agreement with normality, whereas a plot that departs significantly from linearity is evidence that the distribution is not normal (Neter et al. 1996). Residuals from any type of model should also be plotted against variables omitted from the model that might have important effects on the response. For example, plotting residuals against a time variable or location variable can indicate if there is spatial or temporal independence. Correlegrams and bubble are useful plots for identifying lack of special independence (Zuur et al 2009). The Durbin-Watson test for autocorrelation is a good test for autocorrelated residuals where the null hypothesis that values of residuals are not dependent of the magnitude of the residual at the previous time step is to be tested (Mendenhall and Sincich 2004). For a situation in which linear models and generalized linear models are suspected of requiring random effects terms, boxplots are helpful. For example, when observations within sites may be dependent, box plots of standardized residuals grouped by sites are useful (Pinheiro and Bates 2000, Zuur et al. 2009). These boxplots should center around the zero line if a random effects term for site is not needed. The need for a random effects slope term may be assessed by comparing plots of residuals or fitted values from the random effects for slope model with plots of residuals or fitted values from the model without a random effects term for slope when these plots are grouped by site (Zuur et al. 2009). Alternatively, AIC and likelihood ration tests can be used (Pinheiro and Bates 2000).

Section 5 – Simulation – Monte Carlo, Sensitivity Analysis and Fuzzification

Section 5.1 – Monte Carlo

To verify physical microhabitat accuracy and precision error analysis and model validation should be done on sample measurements of physical variables taken over a wide variety of streamflows (Bovee 1996). This data may be difficult or too costly to obtain. Monte Carlo simulation in which data is simulated using expected parameter values may be input to the model for model evaluation (Bovee 1996). Monte Carlo simulation has the advantage that there is no limit to the model hypotheses, sample sizes and data distributions that can be used for model testing.

For example sample size requirements to achieve a given level of accuracy and precision of mode output for a prospective study might be determined by simulating data having variable levels of error in each variance component associated with each stage of model building and fitting the model for several alternative sample sizes until the right combination of error and sample size is found to match the researchers required level of accuracy and precision. Worst case and best case scenarios could be developed.

Section 5.2 - Sensitivity Analysis

Some variables such as temperature may require a large number of independent variables to make model predictions of new temperatures at different locations and times for the same and new independent variable settings. Multiple regression models used to make these predictions can be assessed for adequacy using the model goodness of fit, evaluation and validation measures discussed above for discharge. In addition, in order to evaluate models during the variable selection or calibration stage, Bovee (1996) advocates conducting a sensitivity analyses. A sensitivity analysis is a test of a model in which the value of a single variable or parameter is altered, and the result of the change on the dependent variable is observed. The process can be carried out one variable at a time or in groups if it is thought that interaction effects may be important. In one method the investigator changes the value of each parameter or variable by a fixed percentage during each trial (Fuller 1987, Bovee 1996). Sensitivity analysis provides useful information on model adequacy to all stakeholders in the modeling effort. The effects of errors in each of the variables and parameters on the dependent variable can be assessed. This information permits the researcher to identify sensitive (insensitive) variables (those which have a large (small) influence on the dependent variable temperature) that must be reliably estimated or for which larger errors can be allowed to occur (Fuller 1987, Bovee 1996).

Section 5.3 - Fuzzification

Fukuda (In Press) and Fukuda and Hiramatsu (2008) used a technique known as fuzzification in evaluating alternatives among fish habitat preference models. The effectiveness of the fuzzification in fish habitat modelling was assessed by comparing mean square error and standard deviation of the models, and fluctuation in habitat preference curves evaluated by each model. As a result, the effect of fuzzification appeared as smoother curves and was found to reduce fluctuation in habitat preference curves in proportion to the level of fuzzification. The smooth curves would be appropriate for expressing uncertainty in habitat preference of the fish. Fuzzification is the process of transforming discrete values into grades of membership for the purpose of inclusion into a model training set.

CONCLUSION

Fisheries habitat models serve three major purposes: to predict species occurrences using physical and biological variables, to increase the understanding of species-habitat relationships and to quantify habitat requirements. The use of quantitative statistical models to predict the probable occurrence or distribution of species based on relevant key variables is becoming an increasingly important tool in conservation strategy and fishery management.

This literature review has identified and summarized some of the recent and past contributions to evaluating and validating chinook habitat models found in the literature as well as fisheries habitat modeling journal articles for which the model evaluation techniques used could be considered applicable to chinook salmon habitat modeling. Attention has been given to defining key algorithms and equations where these may not be apparent to the reader.

This survey has indicated that there exists a wide variety of models and modeling frameworks within which researcher have viewed the problem of predicting fish distribution and distribution of fish spawning sites. Equally diverse are the methods used to evaluate and validate these models. Nonparametric methods have gained popularity in recent years. Methods such as cross validation and bootstrapping are appropriate for their lack of distributional assumptions though traditional methods such as hypothesis testing and rank based tests are still commonly applied.

New estimation techniques such as artificial neural networks and fuzzy rule based modeling show promise and have simulation based evaluation and validation components.

Table 2. Index for Chinook salmon habitat model evaluation and validation literature.

	Evaluation/Validation Method	Section	Objective of Test	Page
Parametric	Correlation	1.1	Advantages, Disadvantages, p-values, C.I.'s.	2 - 6
	Pearson CC	1.1	Linear Assoc. between 2 vars., Test multicollinearity, Eval. Input,vars.	
	R2	1.1	GOF, Model comparison, Prop. of variation explained.	
	Multiple R2	1.1	GOF, Model comparison, Prop. of variation explained, resub.	
	R2adj	1.1	GOF,Adjusted for resub. Bias, Model comparison, resub.	
	Multiple CC	1.1	Conditional CC. Between vars. When 1 or more accounted for.	
	r2 - Logistic Reg	1.1	GOF, Correlation coefficients for logistic reg., overall measure of fit.	
	P-values	1.2	Probability of more extreme test statistic under null, Critique of.	7, 8
	t-tests, ANOVA	1.3	Tests for 2 or more popln. means, Test diff. between phys. habitat vars.	8, 9
	Paired t-tests	1.3	Evaluate diff. between model vars., data are paired time or space.	
	t- tests of Reg Coeff	1.3	Determine sig. of regression coeffs. or diff. with fixed value.	
	Mean Squared Error	1.3	Measure residual unexplained error. Compare models.	
	F-tests	1.3	Compare nested models, Overall sig. of a linear model.	
	Likelihood Ratio	1.4	Model comparison using loglikelihood values, uses Chi-sq. test stat.	9
	ANCOVA	1.4	Test reg. slopes when continuous and categ. vars. in model.	9
	Confidence Intervals	1.4	Prob. intervals for popln. parameters where alpha is specified, test sig.	9
	Prediction intervals	1.4	Prob. intervals to predict a new value of dependent var.	9
	CI's for Reg Coeff	1.4	Prob. Intervals for reg. coeffs., Test sig. of param., Eval. effect size.	9
	AIC, AIC_c	1.5	Akaike information, Model selection, comparison, variable ranking.	9, 10
	AIC Confidence Set	1.5	Best subset of models among all in set; based on information theory.	9, 10
	Hosmer-Lemeshow	1.6	Logistic reg. GOF test based on predicted and obs.values, uses Chi-sq.	10-12

	Evaluation/Validation Method	Section, Page	Objective of Test	Page
Non-Parametric	New Data	2.1	Test model results on new set of data, less biased.	12-17
	Resubstitution	2.1	Hold out some data from model fitting on which to test the model pred.	12-17
	Jacknife - LOO CV	2.1	Leave-one-out CV, performance measure can be pred. errors.	12-17
	K-fold CV	2.1	Test model predictability on K subsets of left out data.	12-17
	Stratified K-fold CV	2.1	Same as K-fold but subsets all have same mean response.	12-17
	Bootstrap SE	2.2	Empirical SE's calculated from replicate samples, (with repl. Resampling).	17-22
	Bootstrap CI's	2.2	Emp.percentiles of resampling statistic, (case sampling, or model based).	17-22
	Bootstrap P.I's	2.2	BS prediction intervals. Evaluate error in predicting new popln. value.	17-22
	Randomization	2.2	Exact tests; Almost any test can be subject of method to obtain p values.	17-22
	Mann Whitney	2.3	2 sample difference in medians; requires symmetric distributions use rank	22
	Kolmogorov Smirnov	2.3	Used to test differences in two distributions.	22
	Chi-squared	2.4	Test independence among counts grouped by categories.	24
	Confusion Matrix	2.4	Error classification matrix; basis of error classification prediction tests.	24
	Cohen's Kappa	2.4	Test rate of correct predictions.	24
	Max Kappa	2.4	Test rate of correct predictions.	24
	ROC Curves	2.4	Test rate of correct predictions.	24
	Spearman CC	2.3	Test rate of correct predictions.	24
Overview				
	Conceptual Basis	3.1	Develop strategy of evaluation/validation based on conceptual model	24-27
	Suitability Index	3.1	Test to determine if HIS is likely to be an index or probability.	24-27

	Evaluation/Validation Method	Section, Page	Objective of Test	Page
Analysis of Residuals	Histogram	4.1	Normality of errors	27, 28
	Normal Probability Plot	4.1	Normality of errors	27, 28
	Run Charts	4.2	Test for temporal independence by looking for runs of pos or neg resids.	27, 28
	Stand.Residual	4.3	Normalize residuals to express as standard deviation units.	27, 28
	Prediction Error	4.3	Measure of degree of accuracy of model predictions.	27, 28
	Fitted vs Observed	4.3	Plots to inspect for places where the model fits or does not fit observed.	27, 28
	Outlier plots	4.3	Plots to determine locations of outlying points.	27, 28
	Fitted vs. Covariates	4.3	Plots to determine model misspecification.	27, 28
Simulation	Monte Carlo	5.1	Examine model properties by testing with simulated data.	28, 29
	Sensitivity Analysis	5.2	Shift parameter values to explore sensitivity of model output to changes.	28, 29
	Fuzzification	5.3	Type of sensitivity analysis.	28, 29

LITERATURE CITED

Agresti, A. 1990. Categorical Data Analysis. John Wiley & Sons.

Ahmadi-Nedushan, B., A. St-Hilaire, M. Berube, E. Robichaud, N. Thiemonge, and B. B. Bobee. 2006. A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Research and Applications 22:503–523.

Bovee, K. D. 1978. The incremental method of assessing habitat potential for coolwater species, with management implications. American Fisheries Society Special Publication 11:340-346.

Boots, B. and A. Getis (1988): *Point pattern analysis*. In: Scientific Geography Series, Volume 8. Sage Publications, Newbury Park, Ca.

Bovee, K. D. 1986. Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology. U. S. Fish and Wildlife Services Biological Report.

Bovee, K.D. 1996 (editor). The Compleat IFIM: A Coursebook for IF 250. U.S. Dept. of the Interior. U.S. Geological Surey, Biological Resources Division. Washington, D.C. 20240.

Bovee, K. D., B. L. Lamb, J. M. Bartholow, C. B. Stalnaker, J. Taylor, and J. Henriksen. 1998. Stream habitat analysis using the instream flow incremental methodology.

Boyce, M.S., Vernier, P.R., Nielson, S.E., Schmiegelow, F.K.A. 2002. Evaluating Resource Selection Functions. Ecological Modelling 157, 281-300.

Breiman, L., Friedman, J.H., Olshen, R., and Stone, C.J. 1984. Classification and regression trees. Wadsworth, Belmont, Calif.

Breiman, L., Spector, P. 1992. Submodel Selection and Evaluation in Regression. The X-Random Case. International Statistical Review. 60, 3, pp. 291-319.

Bovee, K. D., B. L. Lamb, J. M. Bartholow, C. B. Stalnaker, J. Taylor, and J. Henriksen. 1998. Stream habitat analysis using the instream flow incremental methodology.

Burnham, K.P., Anderson, D.R. 2002. Model Selection and Multimodel Inference. Springer.

Castleberry, D. T., J. Cech, J. J. , D. C. Erman, D. Hankin, M. Healey, G. M. Kondolf, M. Mangel, M. Mohr, P. B. Moyle, J. Nielsen, T. P. Speed, and J. G. Williams. 1996. Uncertainty and instream flow standards. Fisheries 21:20-21.

Conover, W.J. 1980. Practical Nonparametric Statistics. John Wiley & Sons.

Davison, A.C., Hinkley, D.V. 1997. Bootstrap Methods and their Application. Cambridge.

Efron, B. 1982. The Jacknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics.

Efron, B., and Tibshirani, R. 1991. Statistical analysis in the computer age. Science (Wshington, D.C.), 253: 390-395.

Efron, B., and R. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38-49.

Freeman, E. A., and G. Moisen. 2008. PresenceAbsence: An R Package for Presence Absence Analysis. Journal of Statistical Software 23:1-31.

Fukuda, S. In Press. Consideration of fuzziness: Is it necessary in modelling fish habitat preference of Japanese medaka (Oryzias latipes)? Ecological Modelling In Press, Corrected Proof.

Fukuda, S., Hiramatsu, K. 2008. Prediction ability and sensitivity of artificial intelligence-based habitat preference models for predicting spatial distribution of Japanese medaka (Oryzias latipes). Ecological Modelling 215, 301-313.

Fuller, W.A. 1987. Measurement Error Models. Wiley.

Gallagher, S.P. and M. F. Gard. 1999. Relationship between chinook salmon (Oncorhynchus tshawytscha) redd densities and PHABSIM-predicted habitat in the Merced and Lower American rivers, California. Can. J. Fish. Aquat. Sci. 56: 570-577.

Geist, D. R., J. Jones, C. J. Murray, and D. D. Dauble. 2000. Suitability criteria analyzed at the spatial scale of redd clusters improved estimates of fall chinook salmon (Oncorhynchus tshawytscha) spawning habitat use in the Hanford Reach, Columbia River. Canadian Journal of Fisheries and Aquatic Science 57:1636-1646.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993-1009.

Hirzel, A. H., G. L. Lay, V. Helfer, C. Randin, and A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling 199:142-152.

Hosmer, D.W., Lemeshow, S. 2000. *Applied Logistic Regression*. Second Edition. John Wiley & Sons.

Huberty, C.J and Olejnik, S. (2006) *Applied MANOVA and Discriminant Analysis*. New York, Wiley & Sons.

JOHNSON, D.H. (1999). "The insignificance of statistical significance testing, " Journal of Wildife
Management, 63, 763-772.

Johnson, R.A., Wichern, D.W. 2002. Applied Multivariate Statistics. Prentice Hall.

Knapp, R. A., and H. K. Preisler. 1999. Is it possible to predict habitat use by spawning salmonids? A test using California golden trout (*Oncorhynchus mykiss aguabonita*). Canadian Journal of Fisheries and Aquatic Science 56:1576-1584.

Kohavi, R. 1995. "A study of cross-validation and bootstrap for accuracy estimation and model selection." Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, edited by C.S. Mellish, pp. 1137-1143. San Francisco, CA: Morgan Kaufmann. Available at http:// robotics.stanford.edu/~ronnyk.

Kondolf, G.M., Larsen, E.W., Williams, J.G. 2000. Measuring and Modeling the Hydraulic Environment for Assessing Instream Flows. North American Journal of Fisheries Management 20: 1016-1028.

Leclerc, M., P. Boudreau, J. A. Bechara, and L. Belzile. 1996. Numerical method for modelling spawning habitat dynamics of landlocked salmon, *Salmo salar*. Regulated Rivers: Research & Management 12:273-285.

Leclerc, M., A. Boudreault, J. A. Bechara, and G. Corfa. 1995. Two-Dimensional Hydrodynamic Modeling: A Neglected Tool in the Instream FlowIncremental Methodology. Transactions of the American Fisheries Society 124:645-662.

Liao, J.G. and McGee, D. 2003. Adjusted Coefficients of Determination for Logistic Regression. The American Statistician. 57:3.

Lipsey, M.W., Wilson,D.B. 2000. Practical Meta-Analysis. Applied Social Research Methods Series Vol. 49.

McCullagh, P., and Nelder, J.A. 1989. *Generalized linear models*. Chapman and Hall, London.

Manly, B.F. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Second Edition.

Manly BFJ. 2001. Statistics for Environmental Science and Management. Chapman & Hall/CRC: Boca Raton, Florida.

McCarthy, S. G., J. J. Duda, J. M. Emlen, G. R. Hodgson, and D. A. Beauchamp. 2009. Linking Habitat Quality with Trophic Performance of Steelhead along Forest Gradients in the

South Fork Trinity River Watershed, California. Transactions of the American Fisheries Society 138:506-521.

McHugh, P., and P. Budy. 2004. Patterns of Spawning Habitat Selection and Suitability for Two Populations of Spring Chinook Salmon, with an Evaluation of Generic versus Site-Specific Suitability Criteria. Transactions of the American Fisheries Society 133:89-97.

Menard, S. 2000. Coefficients of Determination for Multiple Logistic Regression Analysis. *The American Statistician. 54, 1*

Mendenhall, W., Sincich, T. 2003. A Second Course in Statistics - Regression Analysis. Pearson Prentice Hall.

Milhous, R., M. Updike, and D. Snyder. 1989. PHABSIM system reference manual: version 2. U.S. Fish Wildlife Service.

Mittlbock, M., and Schemper, M. (1996). Explained variation for logistic regression. *Statistics in Medicine. 15.*

Montgomery, D. C. 2005. Introduction to Statistical Quality Control. 5th Edition.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. Biometrika, **78**, 691-692.

Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W. 1996. Applied LInear Statistical Models. McGraw-Hill.

Olden, J. D., D. A. Jackson, and P. R. Peres-Neto. 2002. Predictive Models of Fish Species Distributions: A Note on Proper Validation and Chance Predictions. Transactions of the American Fisheries Society 131:329-336.

Palm, D., E. Brannas, and K. Nilsson. 2009. Predicting site-specific overwintering of juvenile brown trout (*Salma trutta*) using a habitat suitability index. Canadian Journal of Fisheries and Aquatic Science 66:540-546.

Peterson, J. T., C. R. Jackson, C. P. Shea, and G. Li. 2009. Development and Evaluation of a Stream Channel Classification for Estimating Fish Responses to Changing Streamflow. Transactions of the American Fisheries Society 138:1123-1137.

Pinheiro, J.C., Bates, D.M. 2000. Mixed-Effects Models in S and S-Plus. Springer

Quinn, J.F. and *Dunham*, A.E., *1983*. On hypothesis testing in ecology and evolution. American Naturalist. 122, pp. 602-617.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rizzo, M.L. 2008. *Statistical Computing with R*. Chapman & Hall/CRC.

Rosenfeld, J. S., E. Raeburn, P. C. Carrier, and R. Johnson. 2008. Effects of Side Channel Structure on Productivity of Floodplain Habitats for Juvenile Coho Salmon. North American Journal of Fisheries Management 28:1108-1119.

Royall, R. 1999. Statistical Evidence: a likelihood paradigm. Chapman & Hall/CRC.

Shtatland, E.S., Moore, S., Barton, M.B. 2000. Why we need an R^2 measure of fit in proc logistic paper 256-25, SAS, http://www2.sas.com/proceedings/sugi25/25/st/25p256.pdf

Thielke, J. 1985. A logistic regression approach for developing suitability-of-use functions for fish habitat. Proceeding of the Symposium on Small Hydropower and fisheries. American Fisheries Society.

Thomas, J. A., and K. D. Bovee. 1993. Application and testing of a procedure to evaluate transferability of habitat suitability criteria. Regulated Rivers: Research & Management 8:285-294.

Ward, D. M., K. H. Nislow, and C. L. Folt. 2009. Increased Population Density and Suppressed Prey Biomass: Relative Impacts on Juvenile Atlantic Salmon Growth. Transactions of the American Fisheries Society 138:135-143

Williams, J.G. 1996. Lost in Space: Minimum Confidence Intervals for Idealized PHABSIM Studies. Transactions of the American fisheries Society 125: 458-465.

Williams, J.G. 2001. Testing Models Used for Instream Assessment. Fisheries, 26, 12.

Williams, J. G. 2009. Lost in space, the sequel: spatial sampling issues with 1-D PHABSIM. River Research and Applications 9999:n/a.

Williams J.G. 2009b. Comment on Gard (2009).

Vardeman, S.B., Jobe, J.M. 1999. Statistical Quality Assurance Methods for Engineers. Wiley.

Vaughan, I. P., and S. J. Ormerod. 2005. The continuing challenges of testing species distribution models. Journal of Applied Ecology 42:720-730.

Wood, S.N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC.

Zuur, A.F, Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. 2009. *Mixed Effects Models and Extensions in Ecology with R. Springer*.

APPENDIX B

Red Bluff Interim Pumping Plant Screens Hydraulic Evaluation

Results From Initial Hydraulic Evaluation Of Cone Screens At Tehama Colusa Canal Authority's Interim Pumping Plant, June 1 – 12, 2009, Red Bluff, California

Team of evaluation participants:

- Steve Thomas, P.E., National Marine Fisheries Service (NMFS)
- Robert Hughes, P.E., California Department of Fish and Game (CDFG)
- Mark Gard, Ph.D., U.S. Fish and Wildlife Service (USFWS)
- Josh Gruber, USFWS

Background

The interim pumping plant was designed and built as a stop gap measure to divert water from the Sacramento River to the Tehama Colusa (TC) Canal in response to an expected mandate calling for delaying "gates in" operation of the Red Bluff Diversion Dam until June 15 annually during the 2009 through 2011 period. Previously, the dam gates were lowered on May 15 of each year to provide a gravity diversion of up to 2,500 cfs to the TC Canal. A new diversion facility, including a new flat-plate fish screen to accommodate water needs of the TC Canal water users without the need for lowering the RBDD has been designed and construction will begin in mid-2010. The new facility will allow normal flows to be supplied to the TC Canal via pumps making the diversion dam obsolete and is expected to be operational in the spring of 2012. The interim project intended to be used in concert with other facilities to provide water to the canal during annual "gates out" operation of the Re Bluff Diversion Dam for three consecutive years until the new long term pumping plant diversion facility is operational.

The interim pumping plant has ten vertical pumps each with a design capacity of 50 cfs. (Figure 1) Pumps are paired to feed five, 36 inch conveyance pipes that lead to the settling basin at the head of the TC Canal. Each pump is screened with a 14 ft diameter conical fish screen manufactured by Intake Screens, Inc (ISI). Each screen has a total surface area of approximately 180 square feet and has a rotating brush cleaning system for debris removal that operates on a programmable timer. Conical screens were developed to operate in tidal and back water areas where water depths are shallow and there is no dominant current in the water body. They were chosen for this project based on the shallow water conditions at the proposed site even though it was doubtful that approach and sweeping velocity criteria could be met with this screen design[1]. A condition of accepting the proposed design was that a hydraulic evaluation would be carried out to determine whether or not the cone screens could be operated in conformance with the State and federal fish screening criteria[2].

[1] NMFS fish screen criteria document, *Fish Screening Criteria for Anadromous Salmonids* (1997) states, "screen design must provide for uniform flow distribution over the surface of the screen, thereby minimizing approach velocity." The CDFG document, *Fish Screening Criteria* (June, 2000) states, "[t]he design of the screen shall distribute the approach velocity uniformly across the face of the screen."
[2] Refer to conditions 6.4 and 6.7 of Incidental Take Permit No. 2081-2009-006-01 issued by the California Department of Fish and Game.

Goal of Hydraulic Evaluation

Goals of fish screen hydraulic evaluations are typically 1) to measure near screen water velocities under a near worst case scenario of diversion rate and river flows expected to be encountered throughout the life of the facility; and 2) to adjust flow control baffles to distribute flow uniformly over the entire screen surface. Give the atypical use of the cone screen technology at the interim pumping plant, there was a third goal to this evaluation: to determine whether or not the cone screens could be operated in conformance with the State and federal fish screening criteria.

Methods

A SonTek 16 MHz Acoustic Doppler Velocimeter (ADV) was used to measure near-screen velocities in three dimensions: X, Y, and Z. The ADV was positioned such that approach velocity was measured directly by the X component of the probe. Sweeping velocities were calculated as the resultant of Y and Z measured values. Raw data for each location were stored in separate files and processed with WinADV, a program developed by the U.S. Bureau of Reclamation. Point-average velocities were processed with Microsoft Excel to produce charts and graphs.

Data were collected on four occasions over a two week period as shown in Table 1. A shallow draft, aluminum boat owned and operated by USFWS was used to provide safe access to the screens. The boat was tied up to structural piles typically within four feet of the top of each screen unit. This distance was thought to provide sufficient buffer against interference with screen velocities.

Table 1. Pumping plant and river data.

Screen #	Date Tested	Pump HP	Pump Pair	Recorded Paired Pumping Rate (cfs)	River Stage at RBDD
1	June 9	300	1 & 2	81.6	239.52
2	June 9	300	1 & 2	81.6	239.52
3	June 9	300	3 & 4	72.7	239.52
4	June 11	300	3 & 4	72.8	239.47
5	June 11	300	5 & 6	77.5	239.47
6	June 11	350	5 & 6	76.6	239.47
7	June 8	400	7 & 8	77.5	239.64
8	June 8	400	7 & 8	77.5	239.64
9	June 11	400	9 & 10	73.0	239.47
10	June 1	300	9 & 10	68.0	239.39

Screen area was divided into forty eight zones in an array of six depths and eight positions (bearings) around each screen unit (Figure 3). Velocity measurements were taken at or near the center of each zone. Positions for each measurement along each bearing and screen area for each zone are shown in Figure 4. ISI manufactured a jig to position the probe that attached to the screens' cleaning systems (Figure 2, Photo 1). By operating the cleaning system and adjusting the jig the ADV could measure near-screen velocities three inches from the screen face at nearly any point on the screen. The probe size prevented measuring velocities within the top two feet on each screen. (Photo 2) Velocity measurements were recorded at a rate of 25Hz for a minimum of 60 seconds.

The original plan called for measuring velocities on all screens under two conditions: 1) with both paired pumps running; and 2) with only one paired pump running. Because two pumps fed each 36 inch conveyance line, the evaluation team theorized that each pump's capacity would vary depending on whether or not the paired pump was also operating. Due to time constraints and Tehama Colusa Canal Authority's water needs, measurements were taken with both pumps operating for all screens except for Screen #10. Initially, both Pumps 9 and 10 were operating, but only three points were measured when Pump #9 was shut down for the remainder of that test.

Results and Analysis

Plots of approach velocity and sweeping velocity data are shown in Appendixes A and B, respectively. Approach velocity data are also presented graphically overlaid on a plan view of the pumping plant in Appendix C.

Approach velocities on Screens 6 – 10 did not exceed 0.45 fps, but only on Screen 8 were approach velocities well distributed over screen all screen area. That said, overall average approach velocities on Screens 7 and 8 were well below the value expected for the measured diversion rate.

Approach velocity distribution on screen numbers 1 – 5 were heavily influenced by the river current. Approach velocities in areas receiving direct impact of the current (i.e. the upstream surface of the screens) far exceeded the design target value. Velocity data indicate water will pass through the porous cones, entering the upstream side and exiting the downstream side.

The steel plate on the upstream side of Screen #1 successfully reduced flow through what would likely otherwise had been the hottest spot on all screens. Approach velocity measurements at bearing 270 degrees were taken directly over the solid plate and ranged from 0.30 to 0.48 fps, despite having a solid barrier three inches away. Approach velocities to either side of the barrier plate at bearings 225 and 315 ranged from 0.07 to 0.62 and 1.37 to 1.90 fps, respectively.

Sweeping velocities varied over a wide range depending on location. On Screen 1, sweeping velocities were 3 – 4 fps on the leading edge, 6 – nearly 14 fps on either side, and approaching 0 fps on the downstream side. Sweeping velocity patterns were similar on Screens 2 and 3, but to a lesser magnitude. All screens had at least one point where sweeping velocity was essentially zero.

Conclusions

Screens located in the main river current (Screens 1 – 3) had hot spots exceeding 1.0 fps, speeds that could present a serious hazard to juvenile salmonids and sturgeon, as well as other fish. Screens 4 and 5 also had hot spots in patterns similar to those on Screens 1 – 3, although to a lesser magnitude.

The overall average approach velocity on Screen #1 was less than zero, indicating more water was exiting the screen than entering it. This clearly was not the case since the pump was operating at the time of the evaluation. The negative average value was likely the result of a too coarse mesh of measurement points for diversion rate calculations purposes. Additional measurement points on screens with large ranges in approach velocity values will improve diversion rate estimates.

The overall average approach velocity values for Screens 7 and 8 were lower than what would have been expected given the measured pumping rate. These data imply the in line flow meter was faulty or there were problems with measuring the approach velocities for these screens. If the actual diversion rate was less than what was measured, approach velocities will be greater and flow distribution may not be as uniform at the full diversion rate than they were when measured during this evaluation.

Only on Screen 8 were approach velocities relatively uniform over all screen area. Adjusting the flow control baffles on Screens 6 – 10 may be appropriate to increase the uniformity of flow distribution over the entire screen surface of those screens.

Adjusting the existing baffles will not likely have much effect on water passing directly through screen units 1 – 5. A completely different baffle system which compartmentalizes screen sections, preventing flow from passing in one side and out the other, would greatly improve approach velocity distribution on screens located in an active current (i.e. Screens 1 – 5).

Sweeping velocity criteria were not always met, especially in the backwater area of Screens 6 – 10. When sweeping velocities are very low screen hot spots accumulate debris and present a greater hazard of impingement than a screen with greater sweeping velocities. In areas where sweeping velocities are very low manual debris removal is important to maintain satisfactory hydraulic conditions.

For most measurement locations, sweeping velocities exceeded approach velocities, in many cases by an order of magnitude or more. At those locations, fish coming in contact with the screen face will likely have sufficient velocity to be deflected off the screen and continue with the prevailing current. In areas where sweeping velocity is low, a screen with hot spots may lead to fish impingement (injury and/or mortality). Turbulence in the vicinity of Screens 1 – 4 may disorient juvenile fish allowing predator species to lie in wait in calmer waters for feeding opportunities.

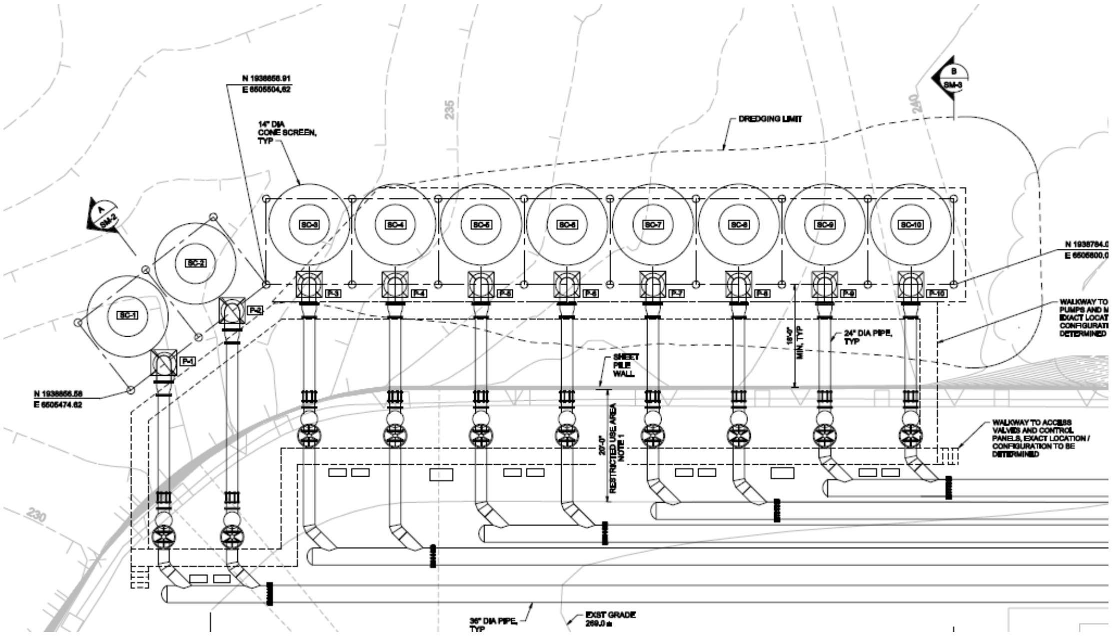


Figure 1. Layout of pumps and screens at the interim pumping plant. Screens and pumps were numbered 1 through 10, left to right.

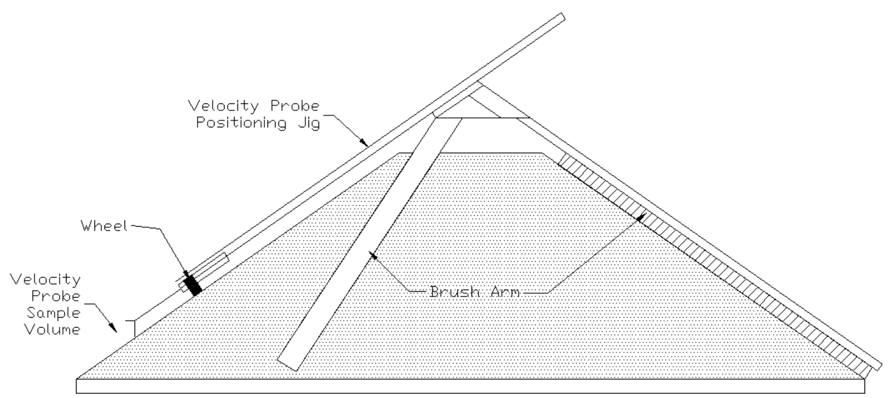


Figure 2. Diagram of equipment used for measuring velocities on cone screens. The jig arm could be raised or lowered to the appropriate elevation on the screen. The jig was attached to the rotating brush system for positioning the velocity probe around the circumference of the screen.

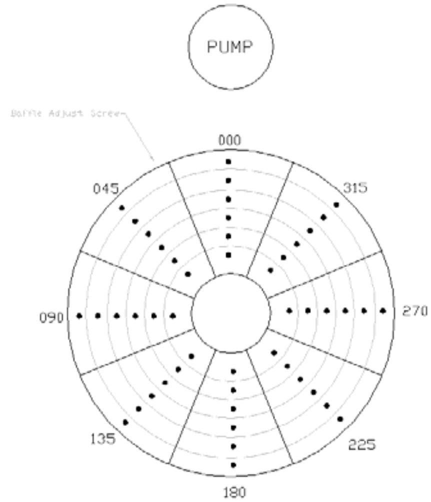


Figure 3. Plan view of locations for velocity measurements on each cone screen: six positions along each of eight bearing angles for a total of 48 measurement locations. The point naming convention used included the bearing angle (with "0" being closest to the pump column), and distance from the toe of the screen (0.5, 1, 2, 3, 4, 5) as shown in Figure 4.

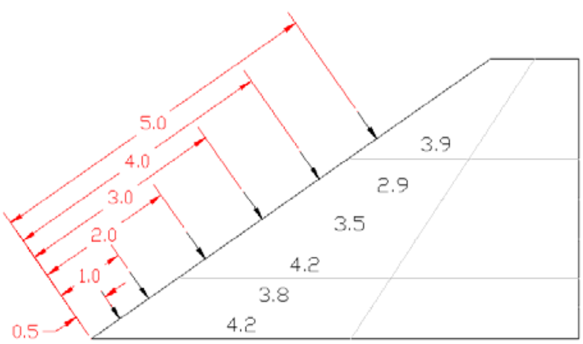


Figure 4. Partial section of a cone screen showing locations where water velocities were measured (arrows, distance values in feet) and the screen zone area associated with those measurements (square feet of screen area per zone). (Zones not shown to scale.)

Photo 1. Mounting the velocity probe and positioning jig to the screen's cleaning system.

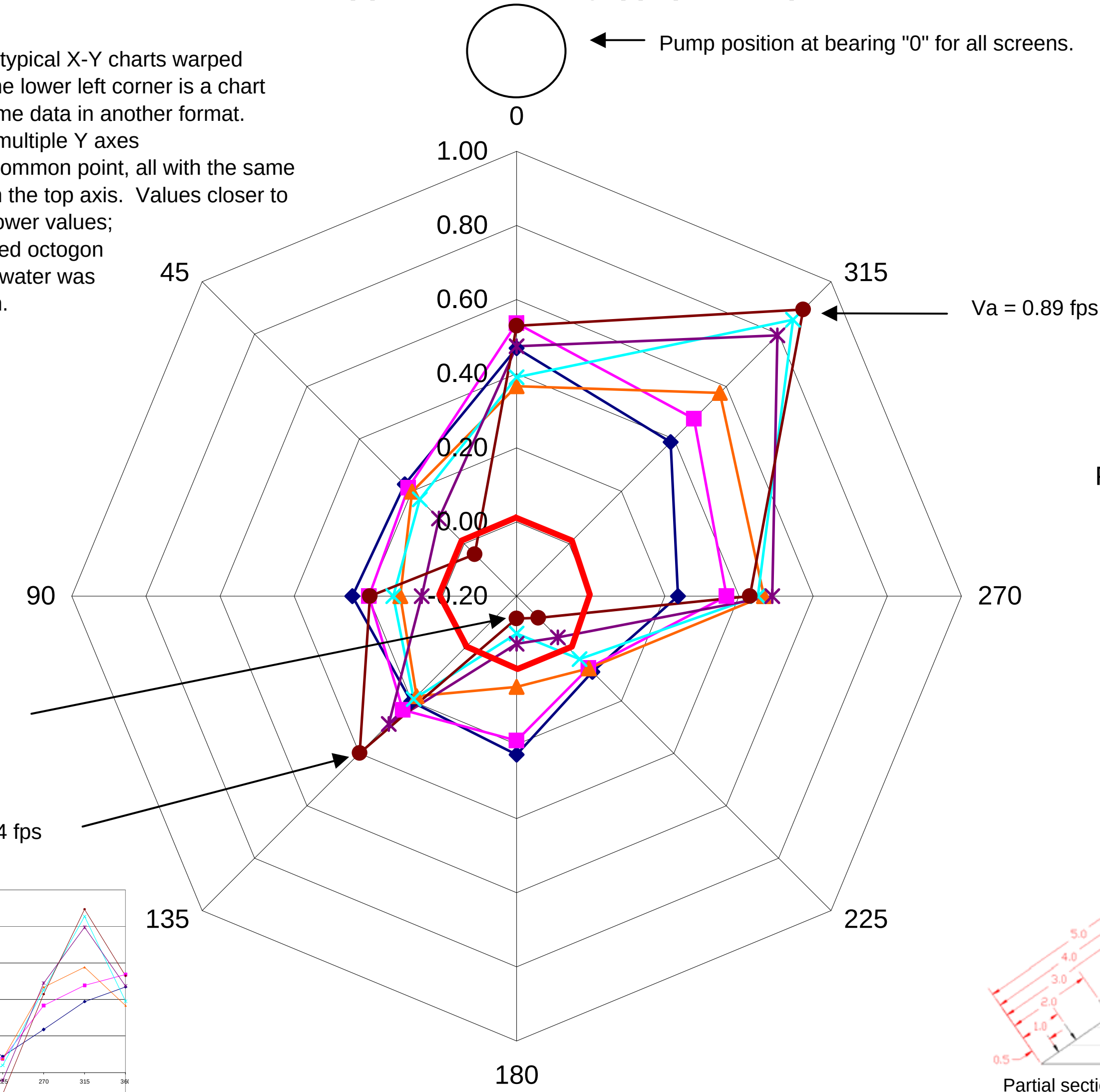
Photo 2. ADV probe in its highest position on the screen measured velocities two feet below the top of the screen panel.

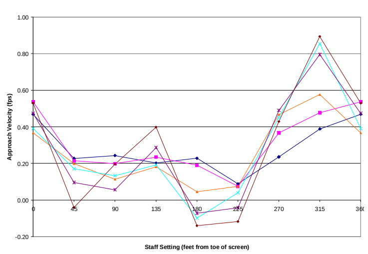
Appendix A

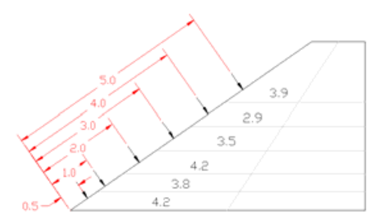
Approach Velocity Plots

Approach Velocity (fps), Example

Radar charts are typical X-Y charts warped into a circle. In the lower left corner is a chart displaying the same data in another format. The charts have multiple Y axes radiating from a common point, all with the same scale as noted on the top axis. Values closer to the center have lower values; those inside the red octogon are negative, i.e. water was exiting the screen.

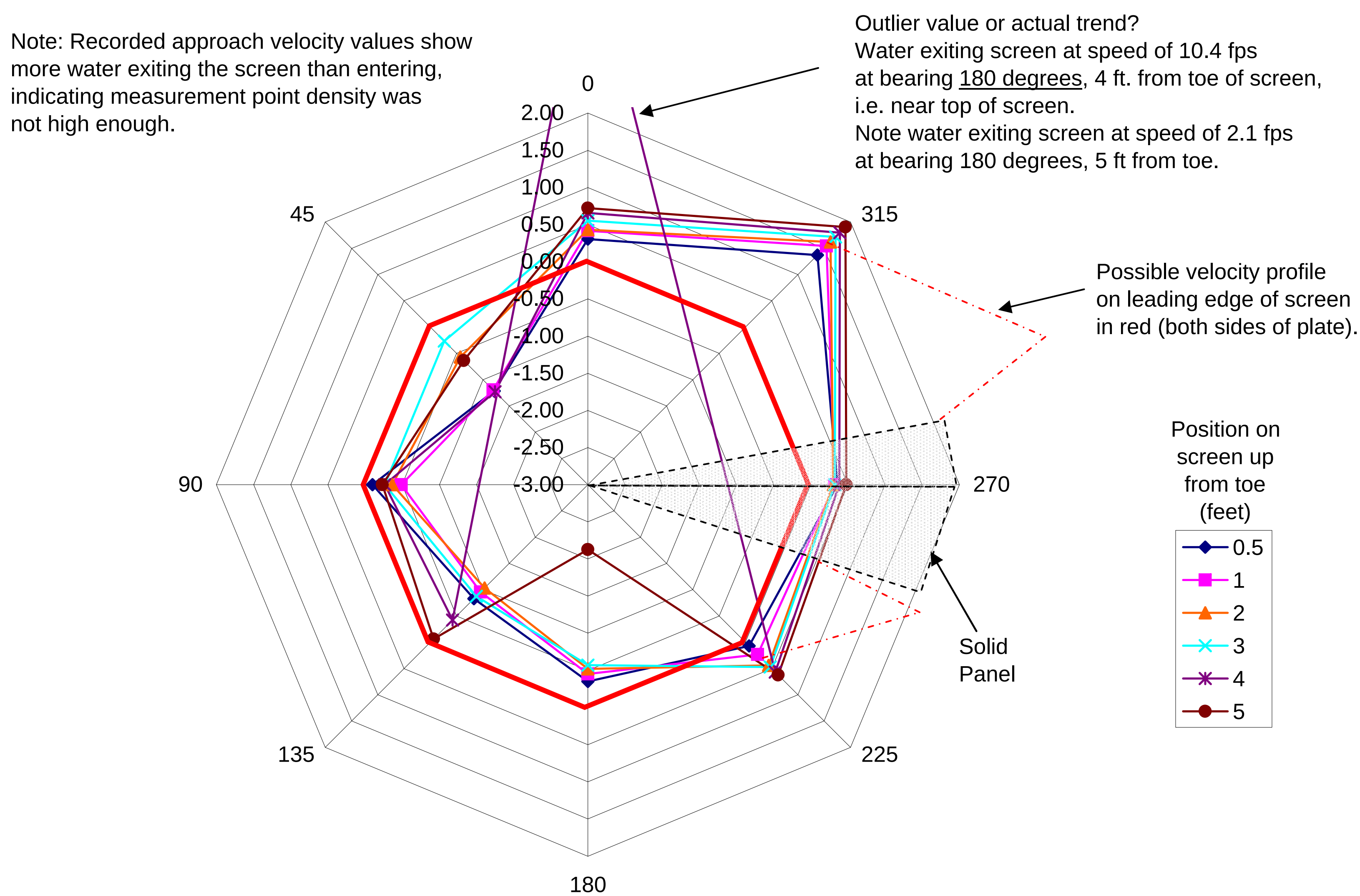


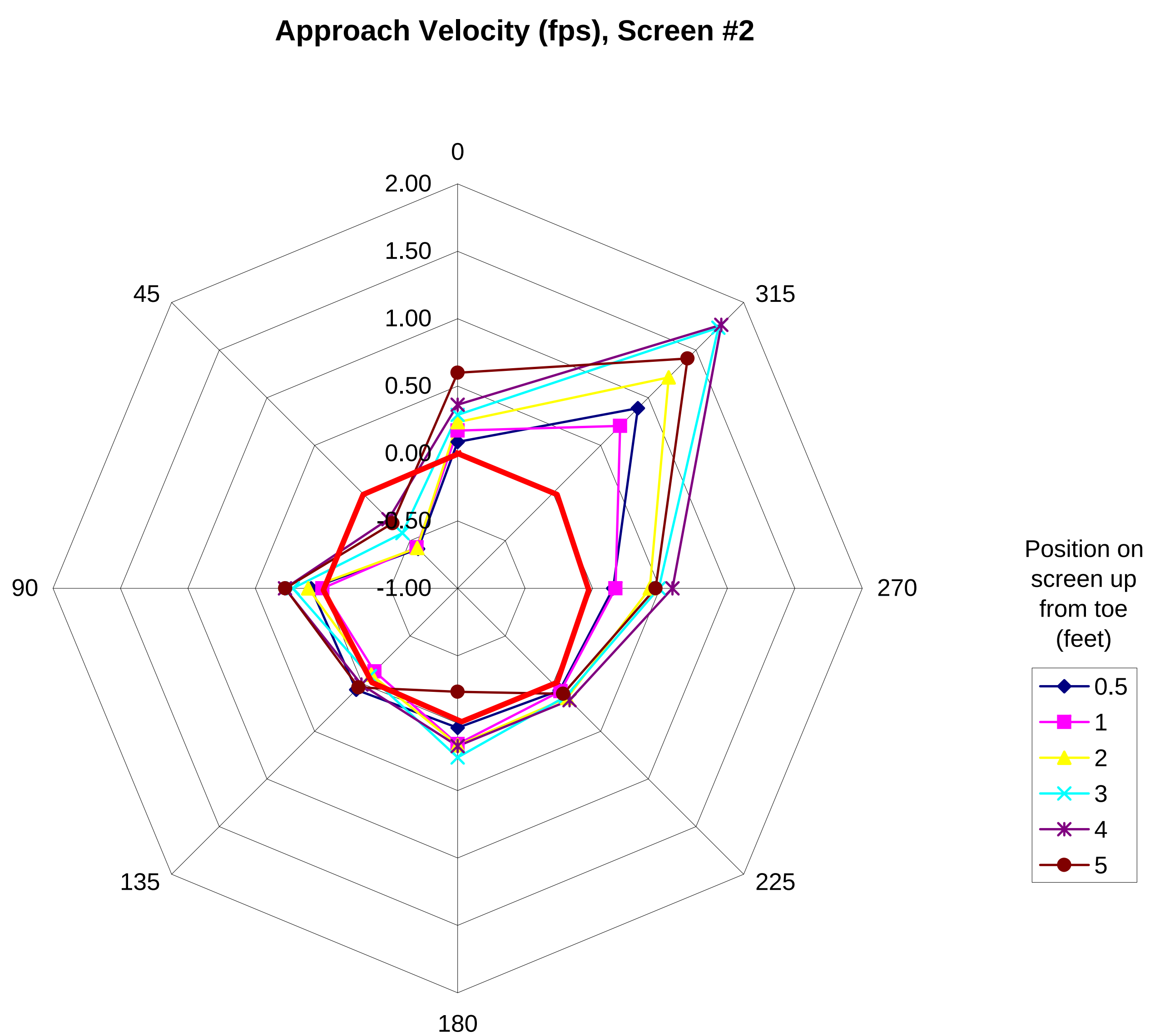




Partial section of a cone screen showing elevations for each measurement.

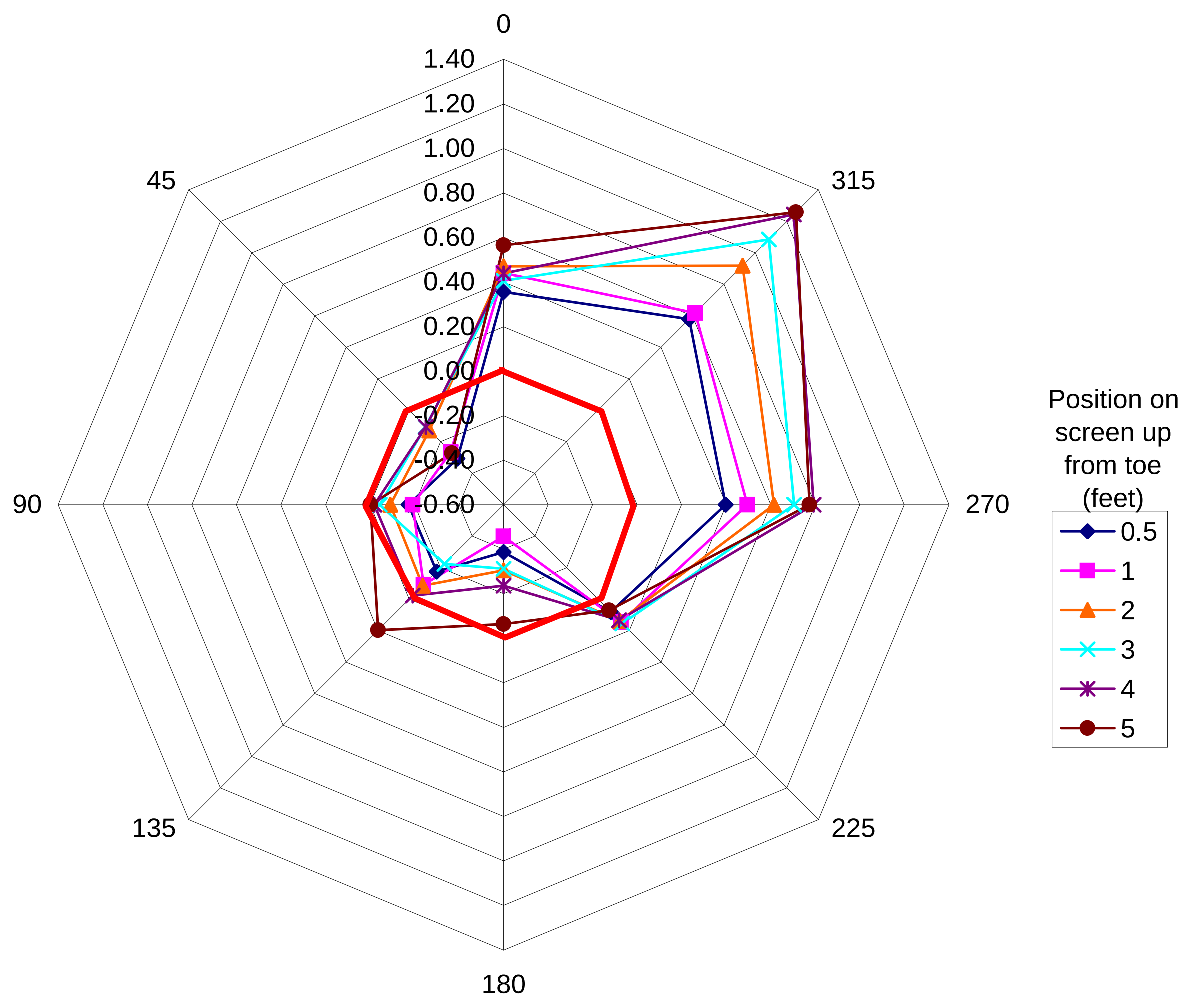
Approach Velocity (fps), Screen #1

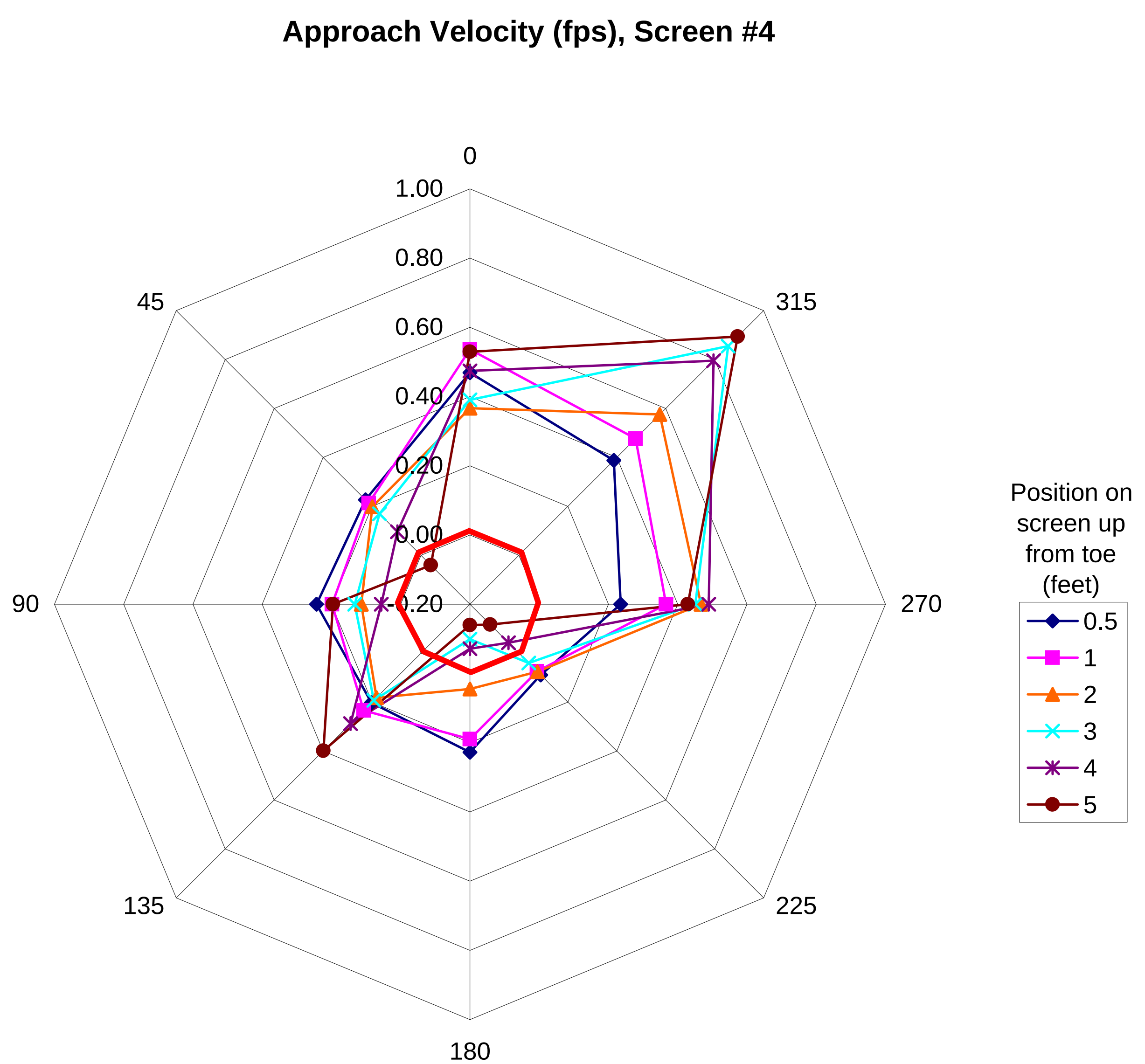


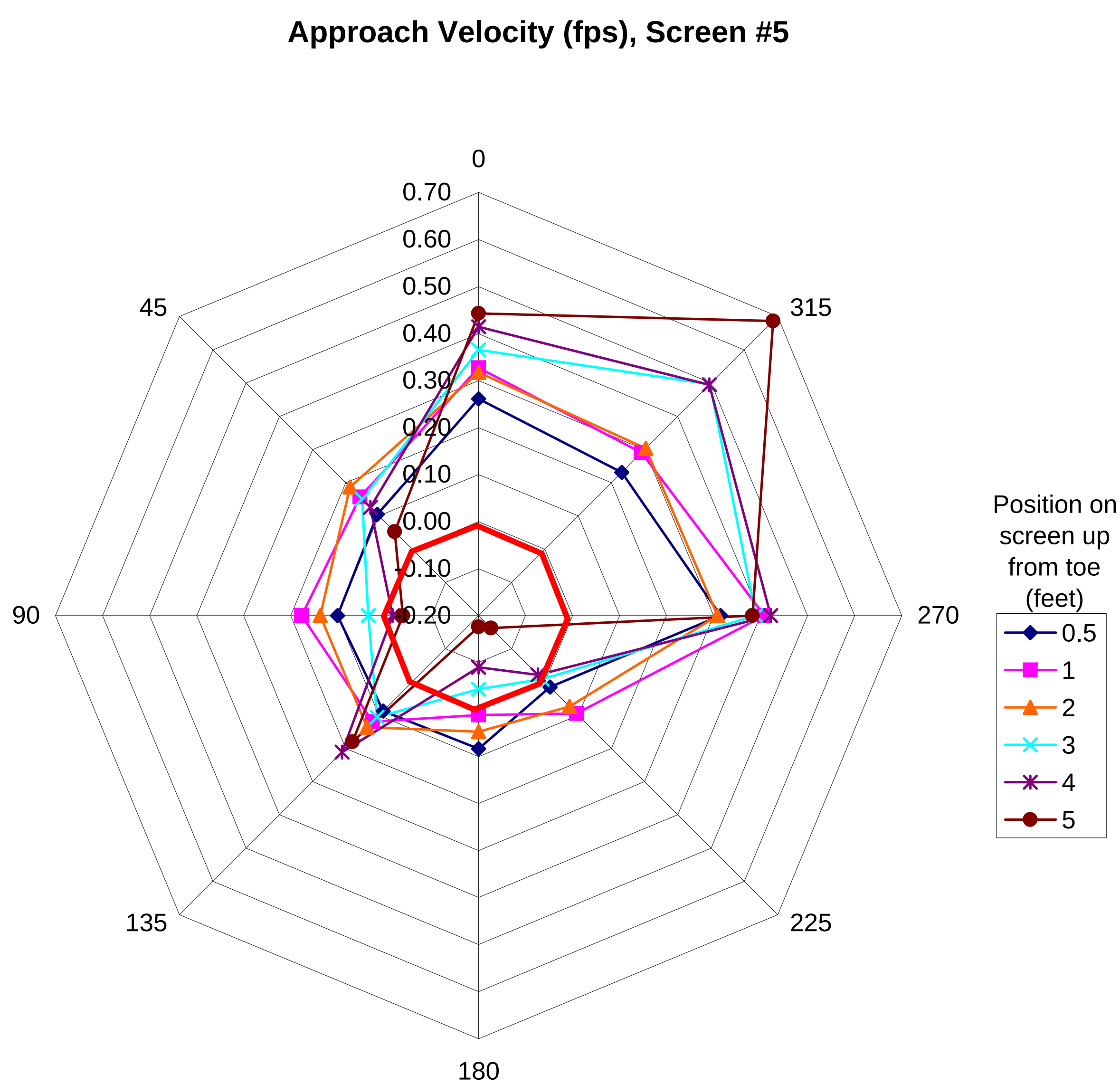


Approach Velocity (fps), Screen #2
0
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
45
315
90
270
135
225
180
Position on screen up from toe (feet)
0.5
1
2
3
4
5

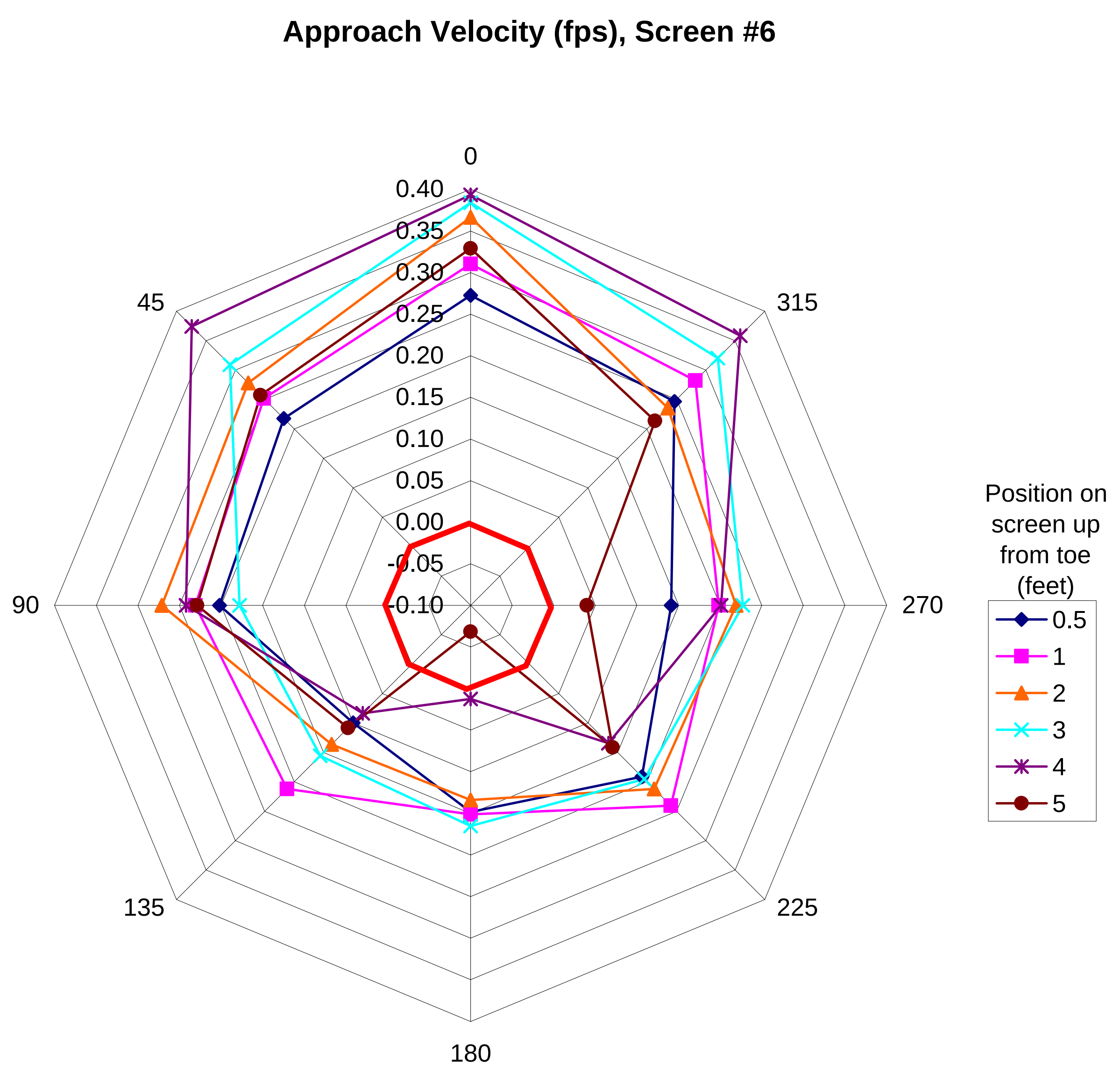
Approach Velocity (fps), Screen #3




Approach Velocity (fps), Screen #4
0
45
90
135
180
225
270
315
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
Position on screen up from toe (feet)
0.5
1
2
3
4
5

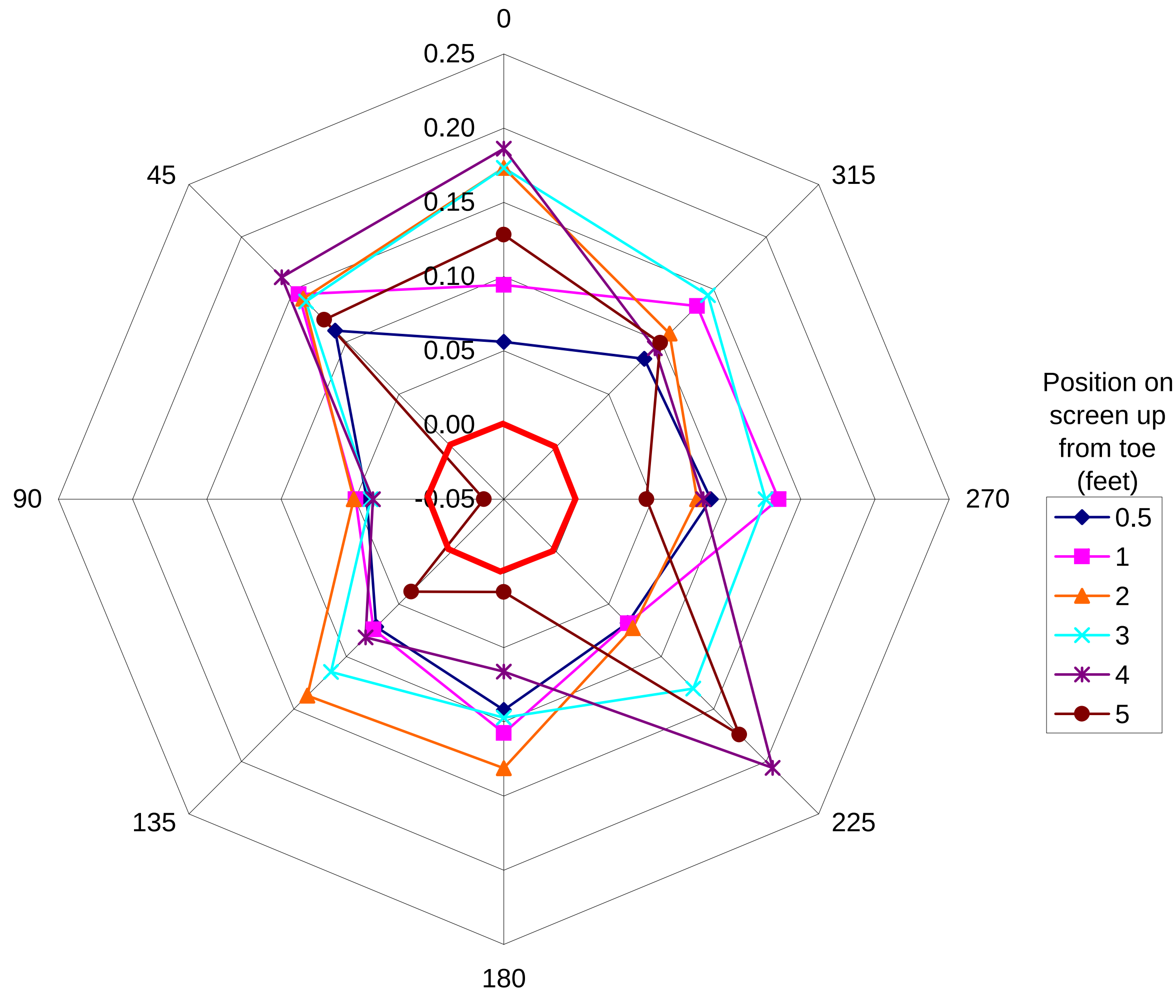


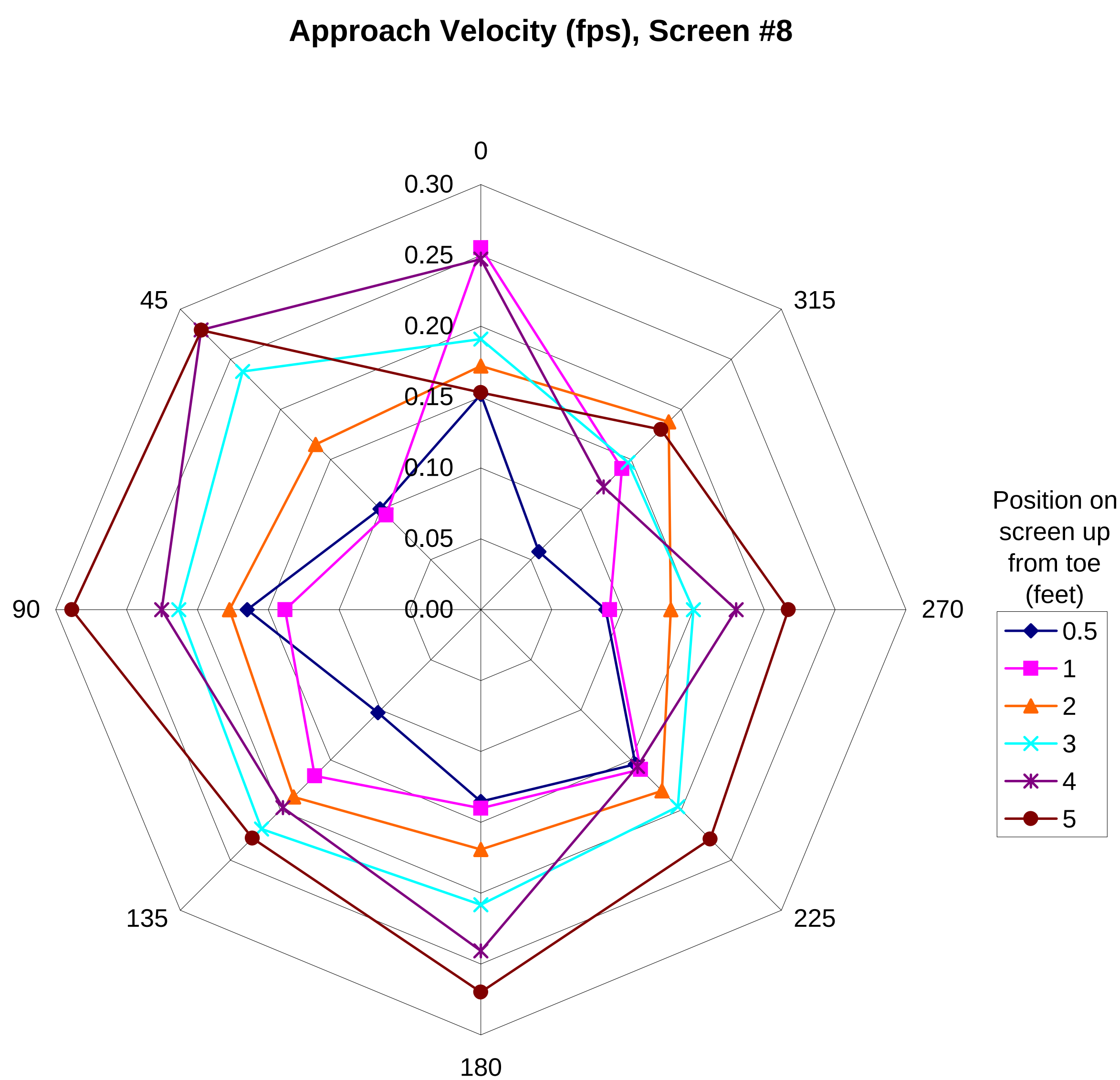
Approach Velocity (fps), Screen #5
0
45
90
135
180
225
270
315
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
Position on screen up from toe (feet)
0.5
1
2
3
4
5

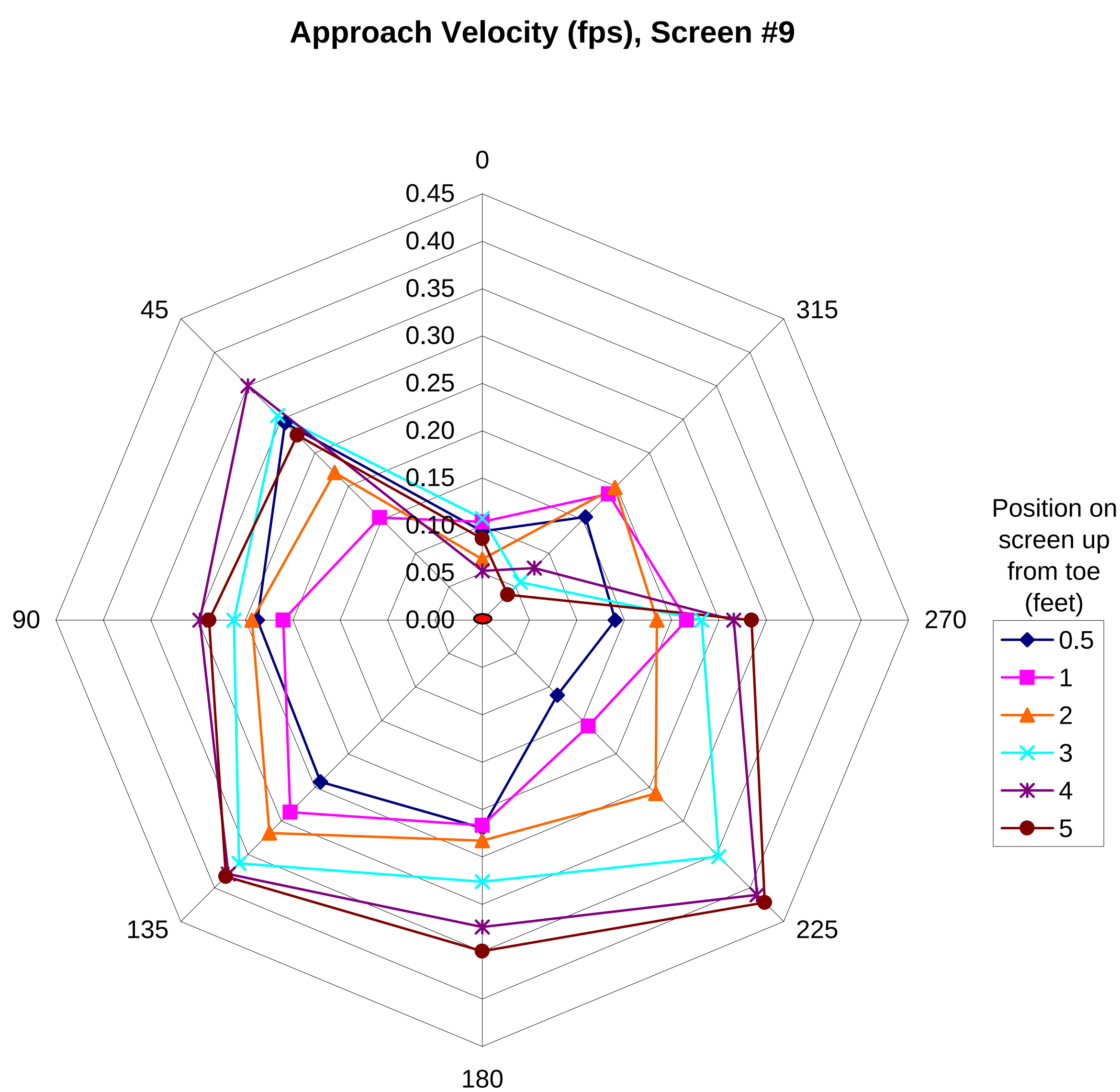


Approach Velocity (fps), Screen #6
0
45
90
135
180
225
270
315
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
Position on screen up from toe (feet)
0.5
1
2
3
4
5

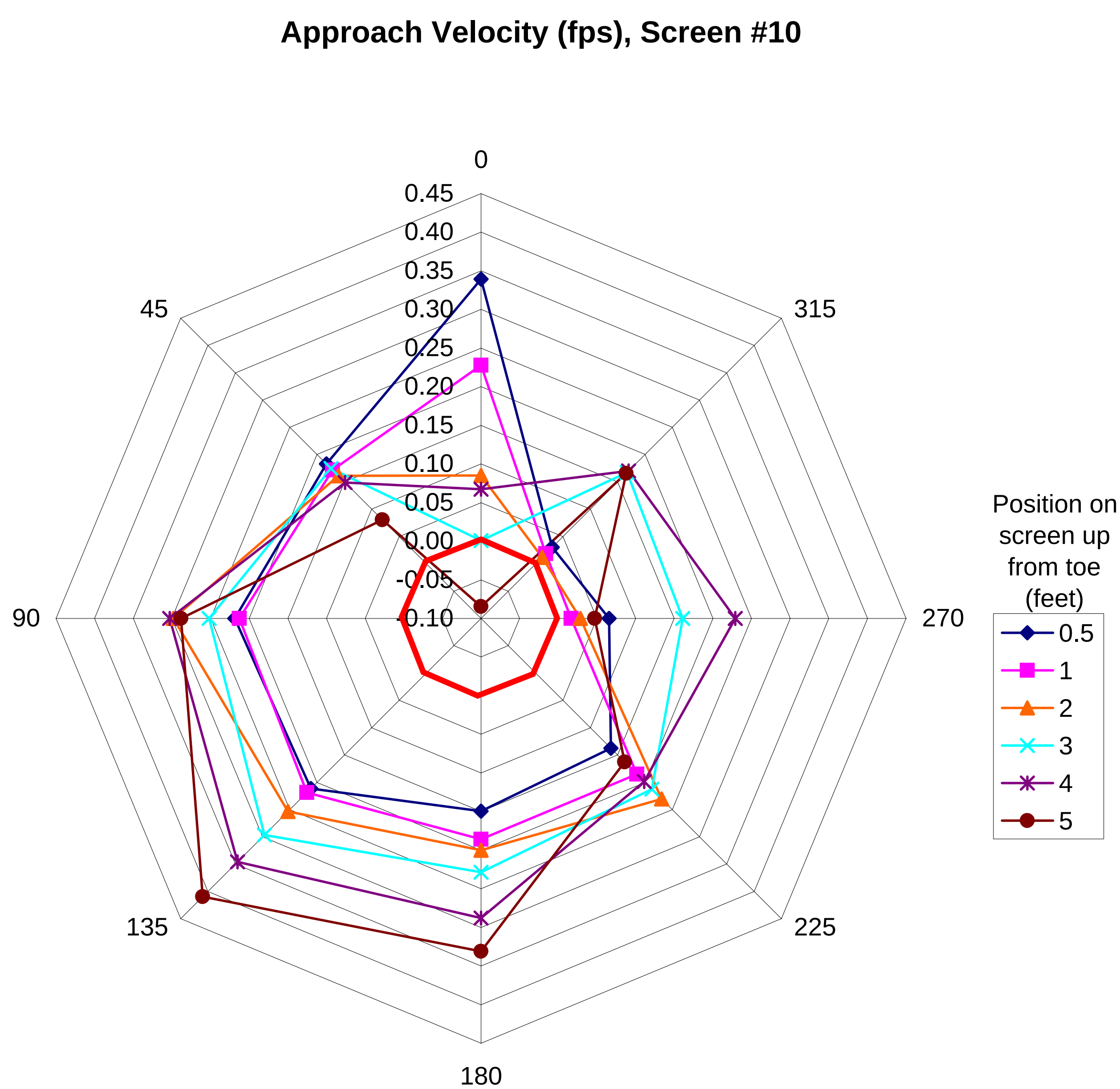
Approach Velocity (fps), Screen #7




Approach Velocity (fps), Screen #8
0
315
270
225
180
135
90
45
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5



Approach Velocity (fps), Screen #9
0
45
90
135
180
225
270
315
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5

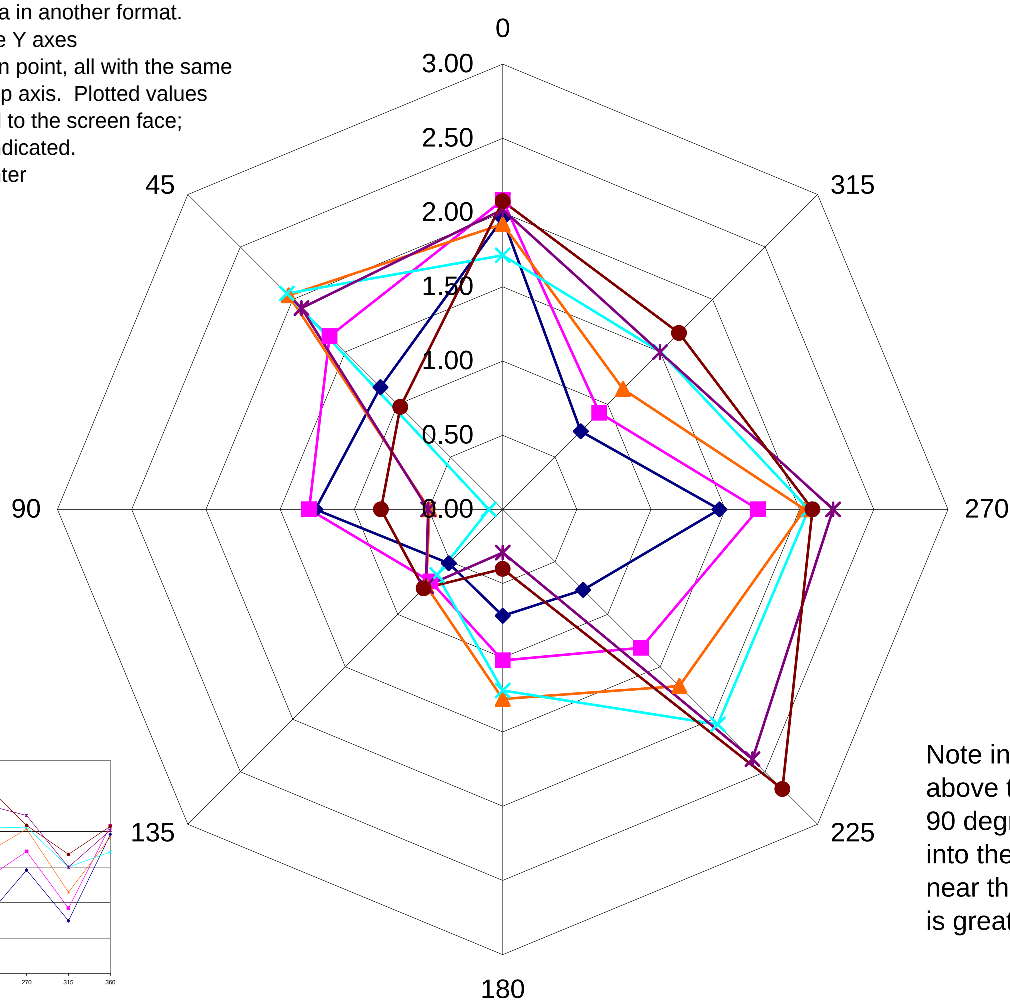

Approach Velocity (fps), Screen #10
0
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
45
90
135
180
225
270
315
Position on screen up from toe (feet)
0.5
1
2
3
4
5

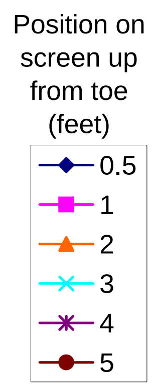
Appendix B

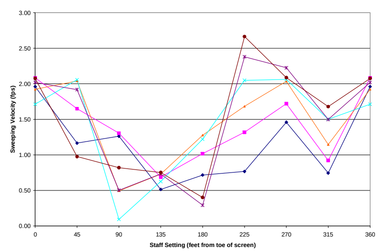
Sweeping Velocity Plots

Sweeping Velocity (fps), Example

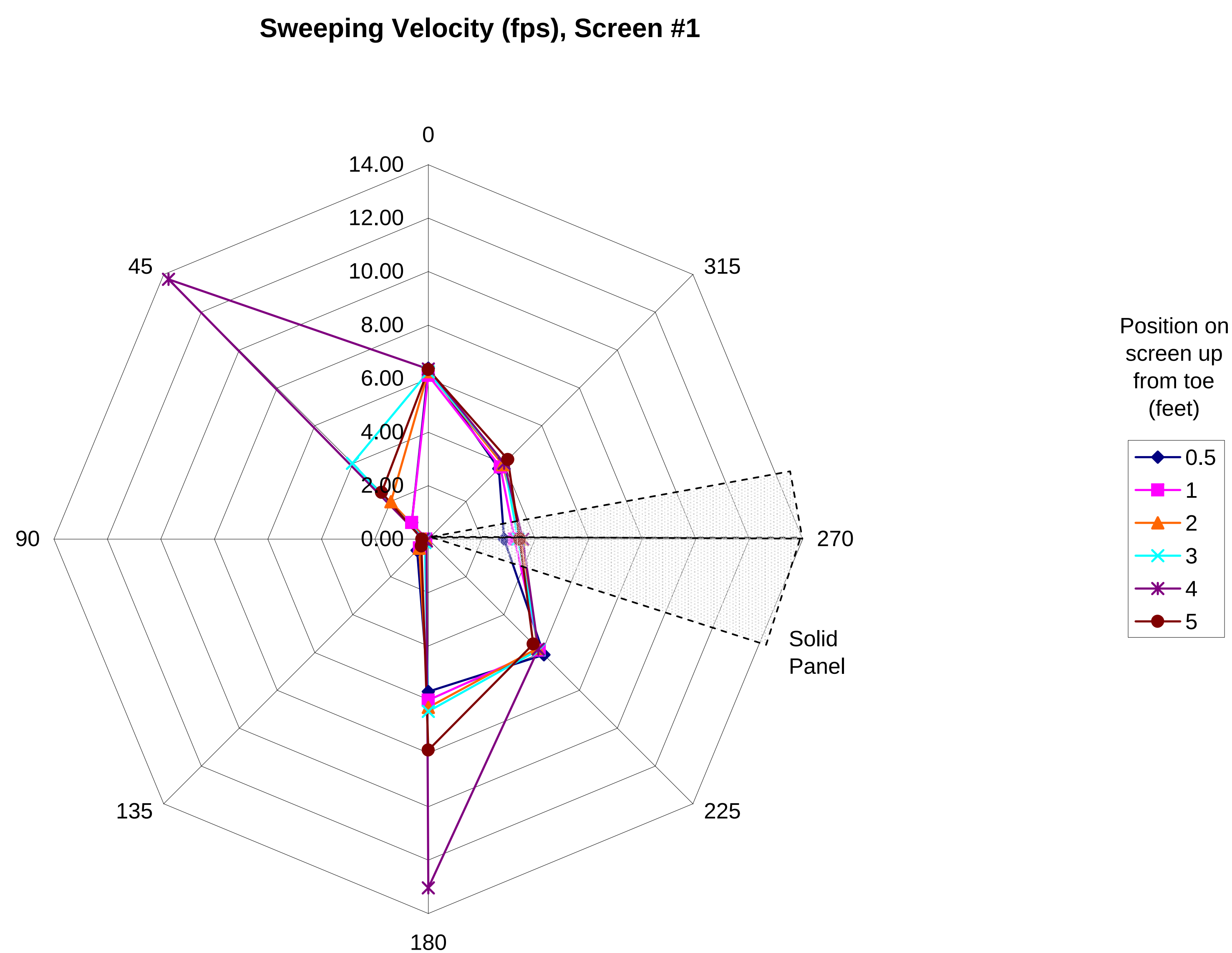
Radar charts are typical X-Y charts warped into a circle. In the lower left corner is a chart displaying the same data in another format. The charts have multiple Y axes radiating from a common point, all with the same scale as noted on the top axis. Plotted values are water speed parallel to the screen face; direction of flow is not indicated. Values closer to the center have lower values.

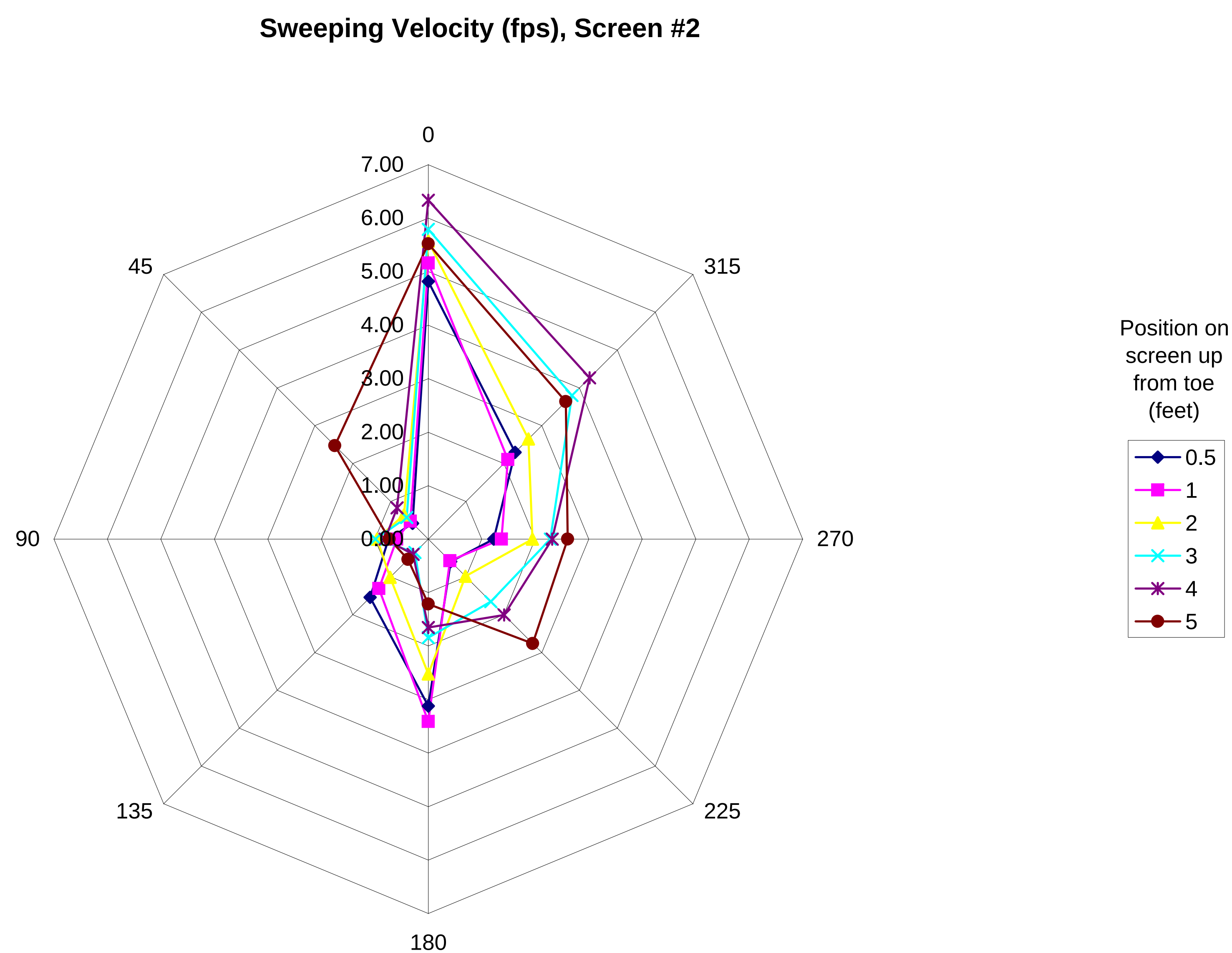


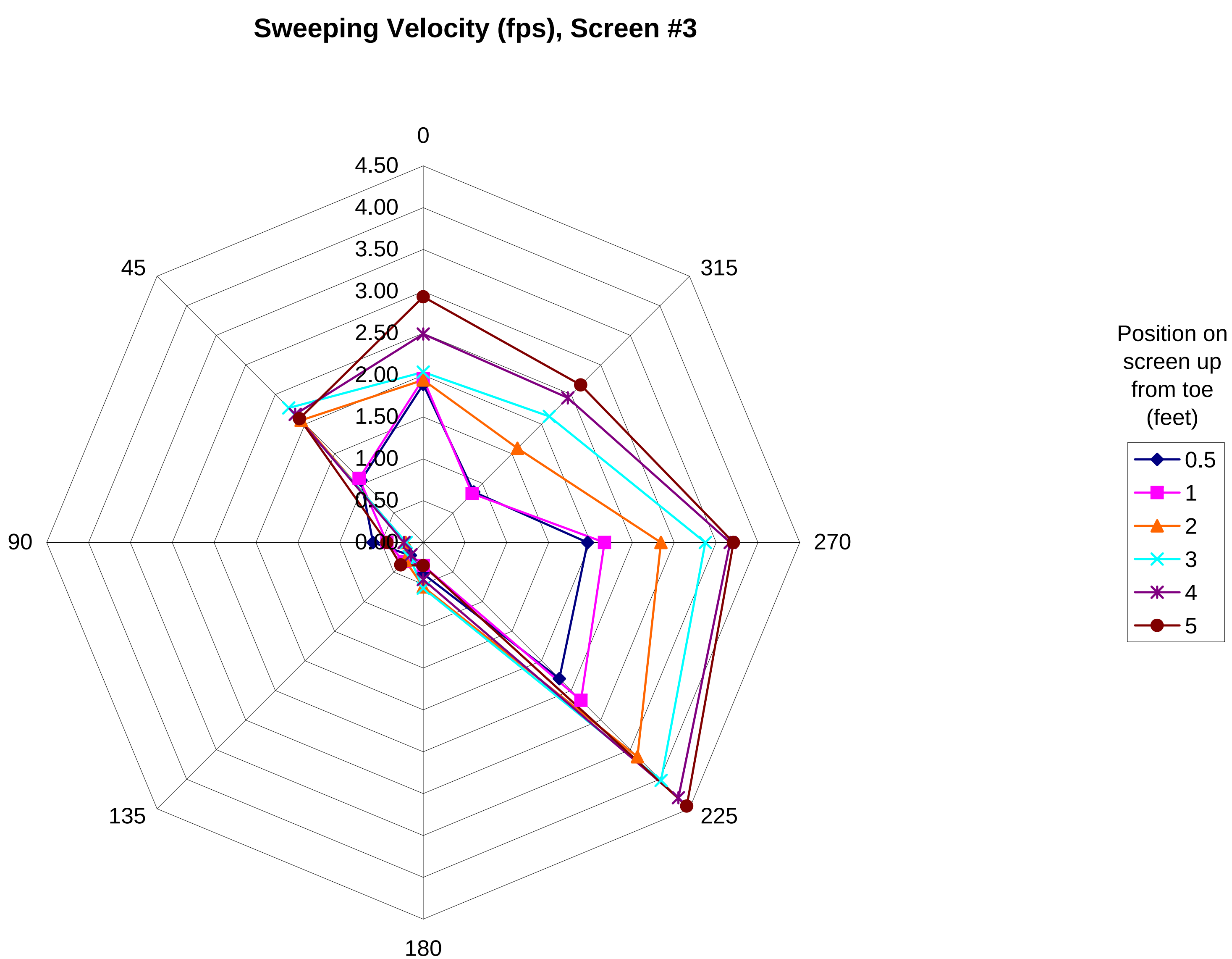




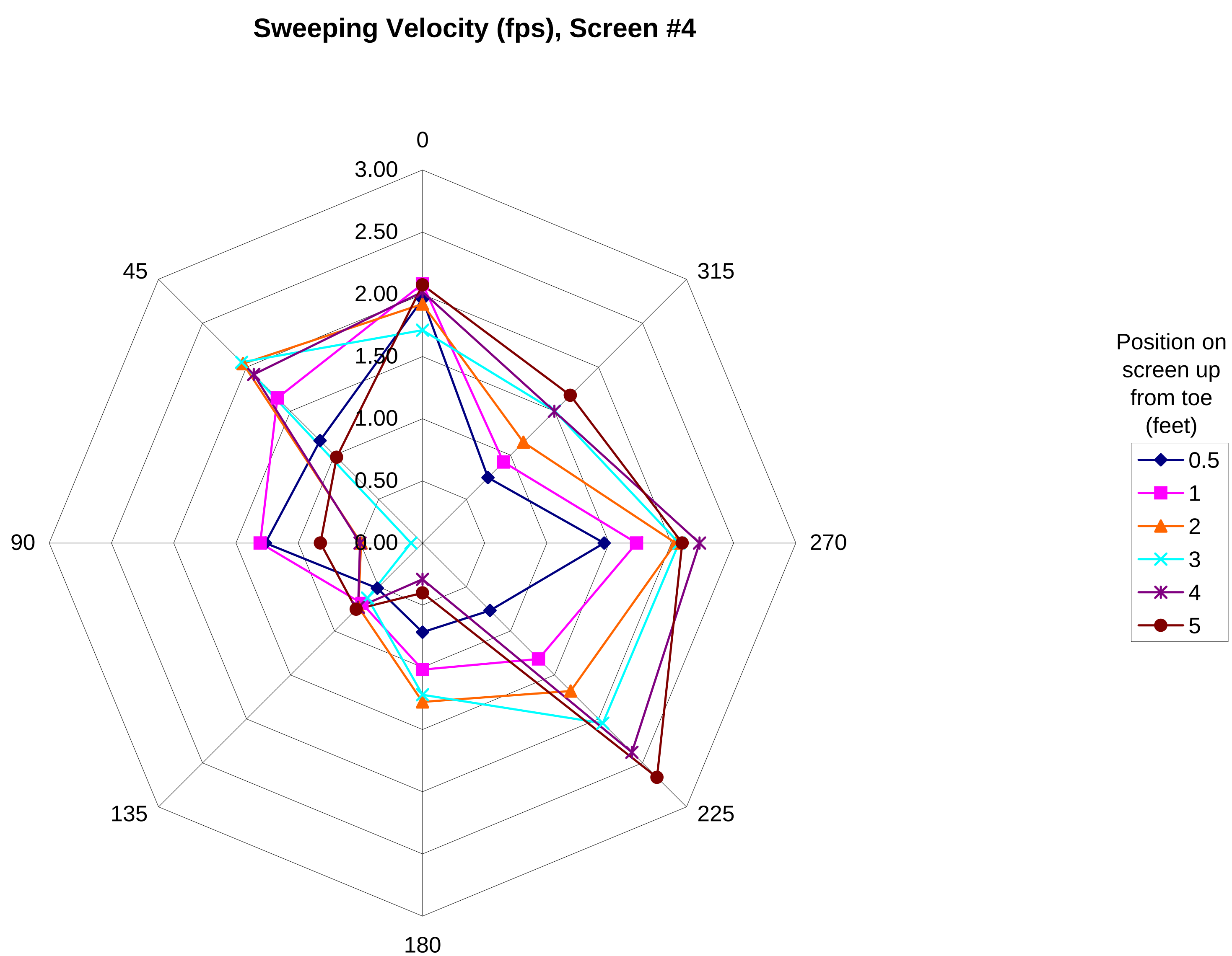
Note in this example, flow at 3 ft above the toe of the screen at bearing 90 degrees is stagnant (or moving into the screen). Sweeping velocity near the surface at bearing 225 degrees is greater than 2.5 fps.


Sweeping Velocity (fps), Screen #1
0
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
45
315
90
270
135
225
180
Solid Panel
Position on screen up from toe (feet)
0.5
1
2
3
4
5

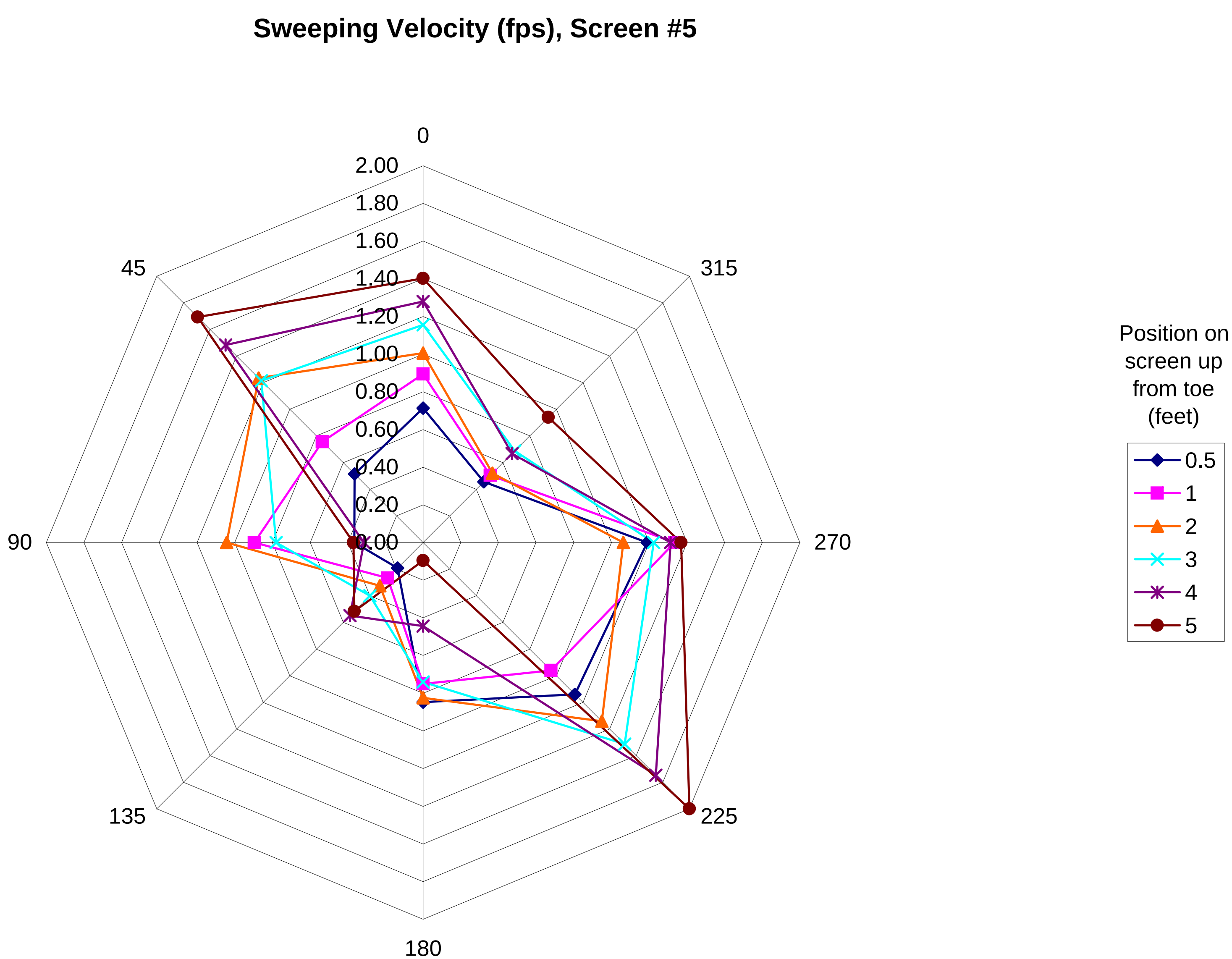

Sweeping Velocity (fps), Screen #2
0
45
90
135
180
225
270
315
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5

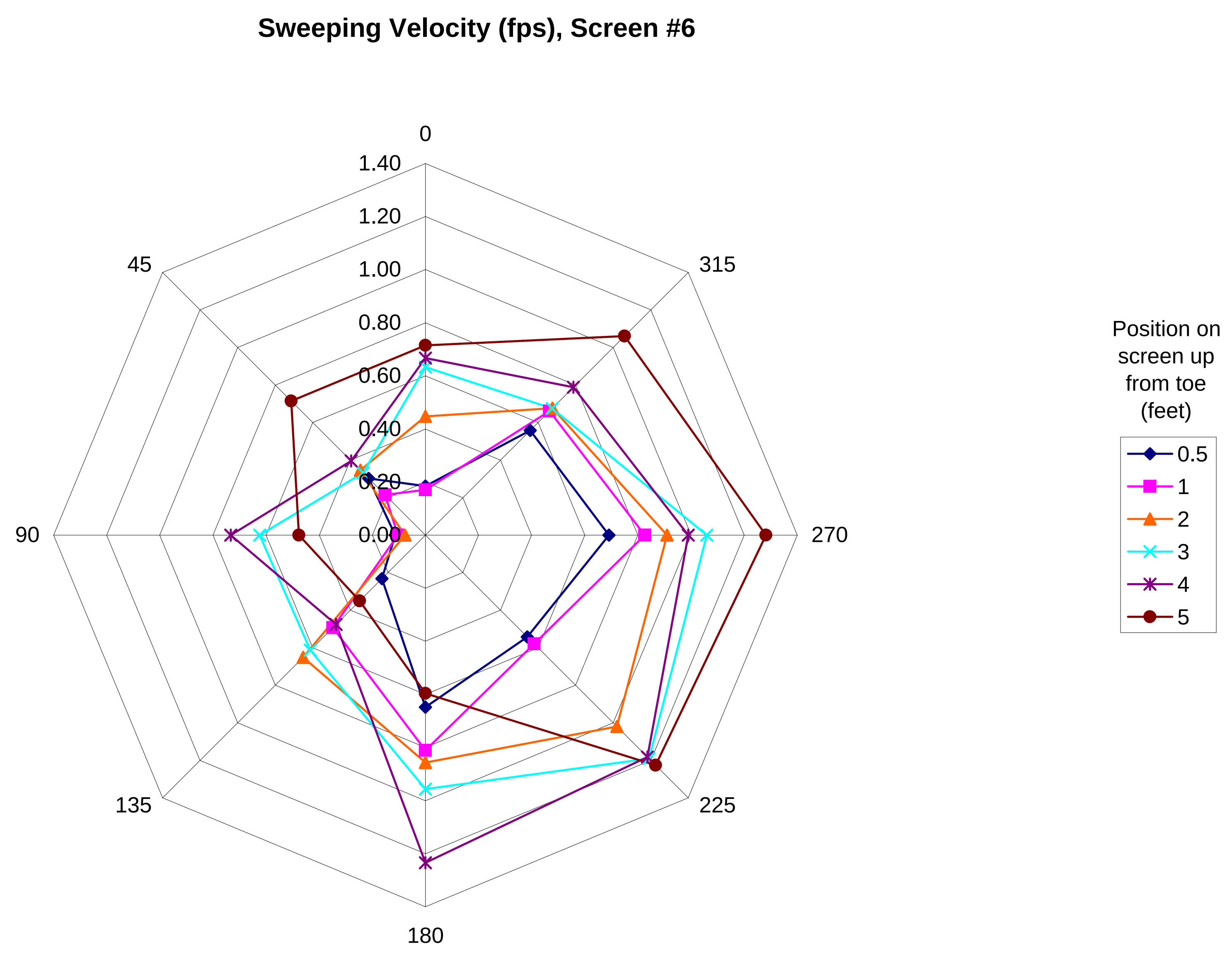


Sweeping Velocity (fps), Screen #3
0
45
90
135
180
225
270
315
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5

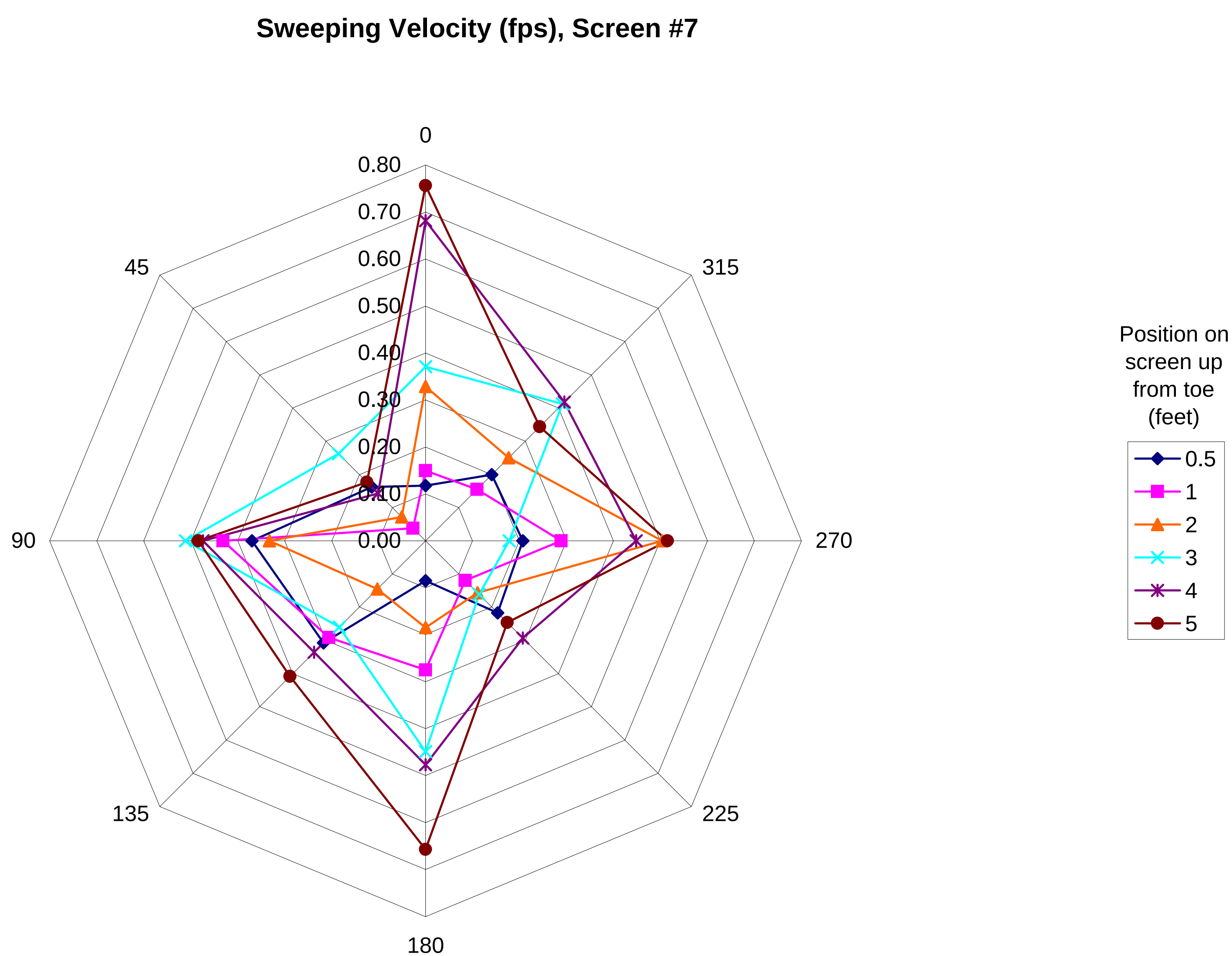


Sweeping Velocity (fps), Screen #4
0
315
270
225
180
135
90
45
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5

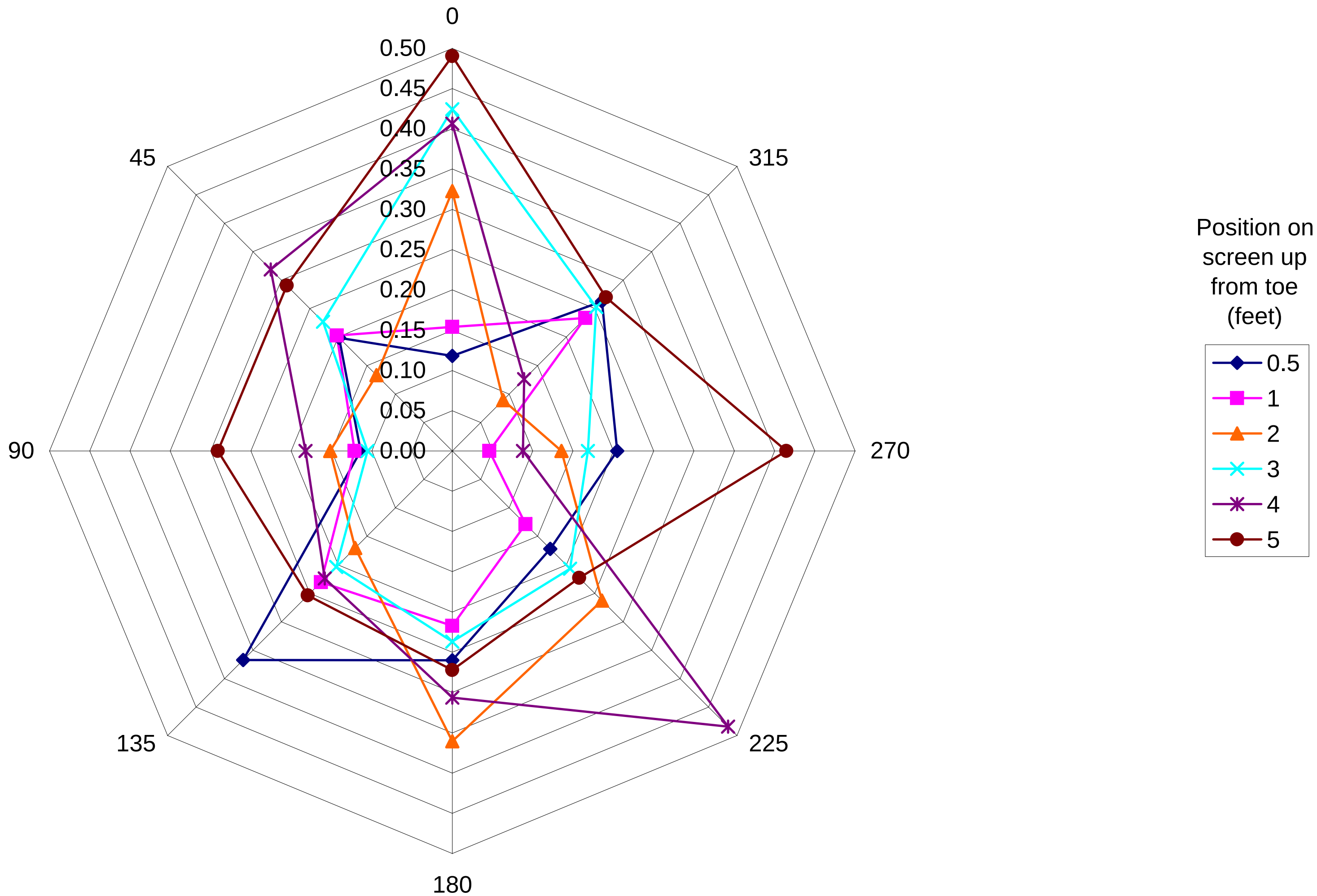

Sweeping Velocity (fps), Screen #5
0
45
90
135
180
225
270
315
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5

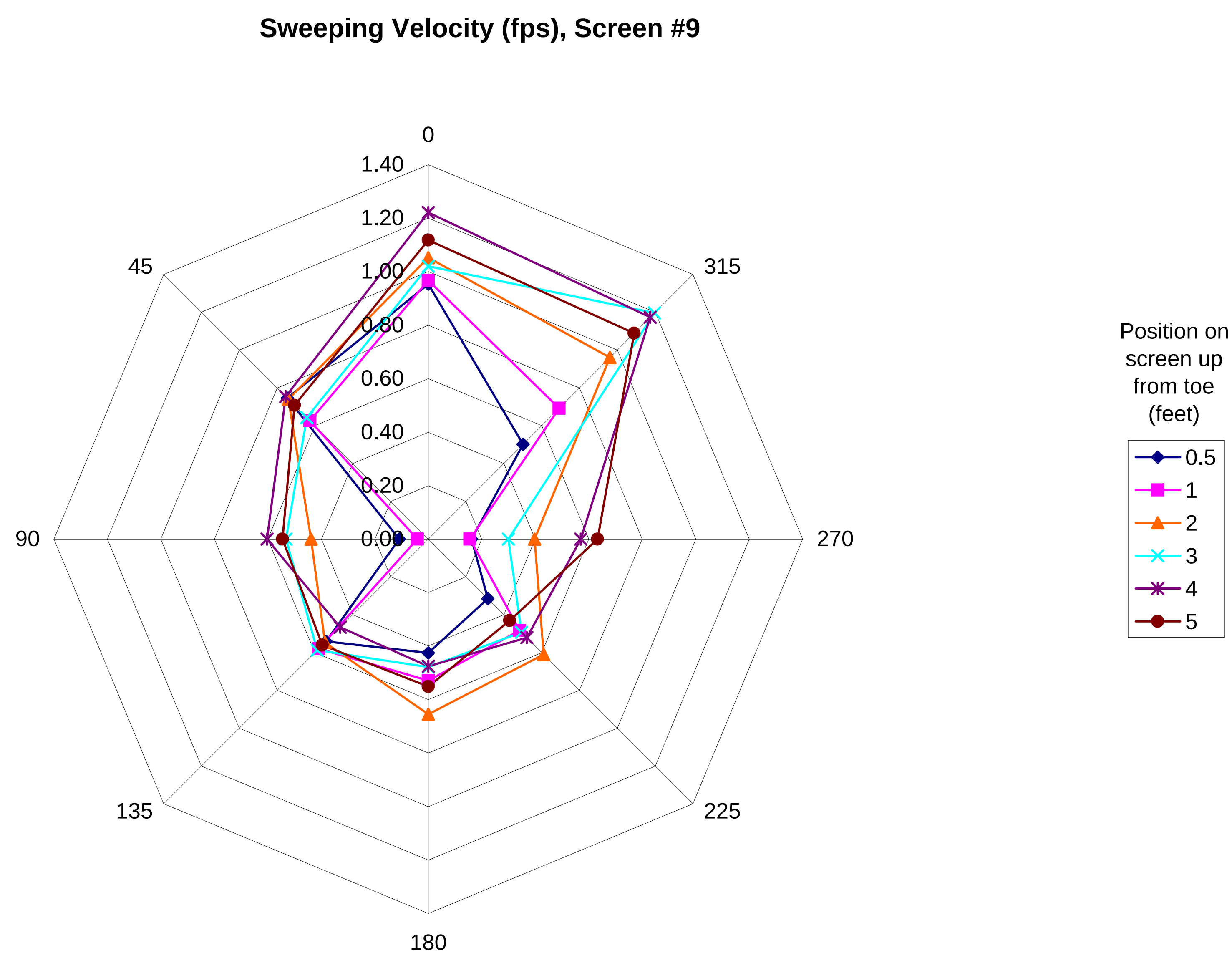


Sweeping Velocity (fps), Screen #6
0
45
90
135
180
225
270
315
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5


Sweeping Velocity (fps), Screen #7
0
315
270
225
180
135
90
45
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5

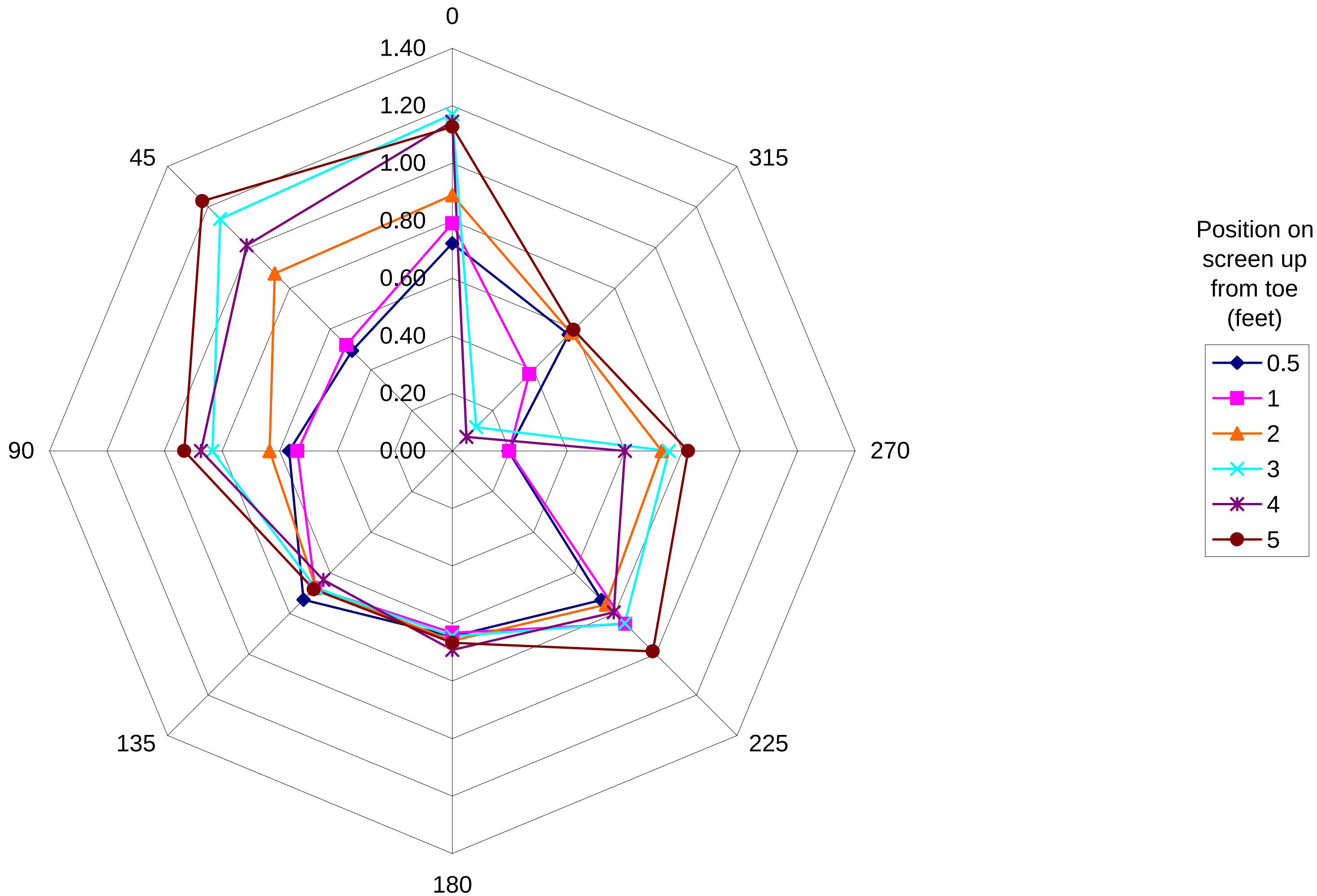
Sweeping Velocity (fps), Screen #8





Sweeping Velocity (fps), Screen #9
0
45
90
135
180
225
270
315
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Position on screen up from toe (feet)
0.5
1
2
3
4
5

Sweeping Velocity (fps), Screen #10

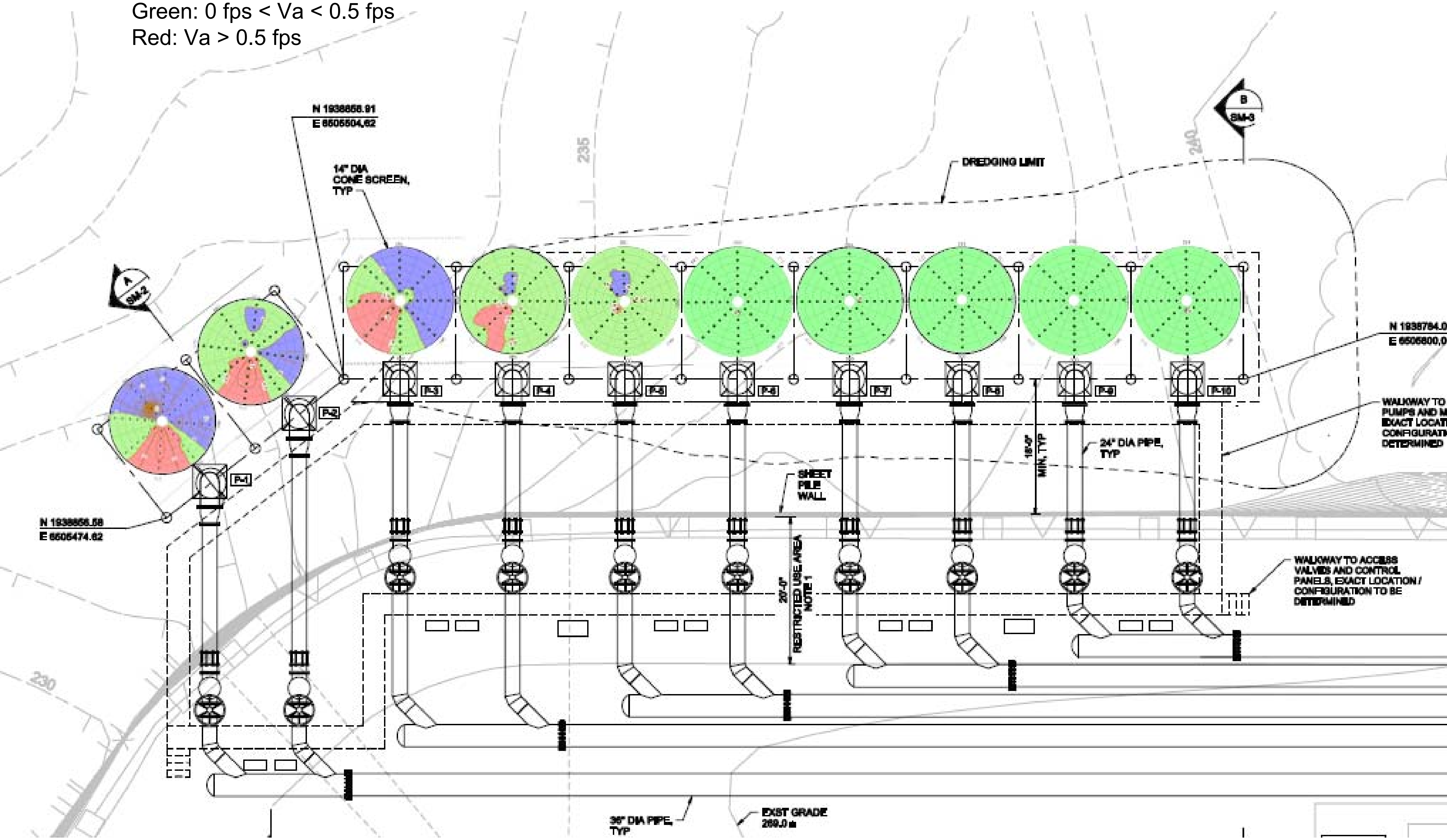


Appendix C

Color Coded Approach Velocity Graphic

Approach velocity data are shown graphically overlaid on a plan view of the pumping plant. Areas of approach velocity greater than 0.5 feet per second (fps) are colored red. Areas with water exiting the screen, i.e. with negative approach velocities, are colored blue. Areas with approach velocity values between 0.0 fps and 0.5 fps are colored green. The design approach velocity criterion for this project was 0.33 fps.

Approach Velocities
Blue: Va < 0 fps
Green: 0 fps < Va < 0.5 fps
Red: Va > 0.5 fps



APPENDIX C

GARD 2009

Intl. J. River Basin Management Vol. 7, No. 1 (2009), pp. 55–71

Comparison of spawning habitat predictions of PHABSIM and River2D models*

MARK GARD, *U.S. Fish and Wildlife Service, 2800 Cottage Way, Room W-2605, Sacramento, CA 95825 USA.*
E-mail: mark_gard@fws.gov

ABSTRACT
This study compared the predictions of two instream flow habitat models, the Physical Habitat Simulation System (PHABSIM) and River2D, with regards to spawning habitat for chinook salmon, *Oncorhynchus tschawytscha*, and steelhead trout, *Oncorhynchus mykiss*. Spawning habitat was simulated with both models for eight sites in the Sacramento River, five sites in the American River and one site in the Merced River, California, using habitat suitability criteria developed from data collected on redds in each of these rivers. For four out of five cases, both models correctly predicted that the combined suitability, calculated as the product of the depth, velocity and substrate suitabilities, of occupied locations was significantly greater than the combined suitability of unoccupied locations. There was little difference in the flow-habitat relationships for each site and set of habitat suitability criteria predicted by the two models. The use of River2D, rather than PHABSIM, is still warranted given its ability to model complex flow conditions which cannot be simulated with PHABSIM.

Keywords: Instream Flow Incremental Methodology; IFIM; chinook salmon (*Oncorhynchus tschawytscha)*; Physical Habitat Simulation system; PHABSIM; Two-dimensional habitat modeling.

1 Introduction

By applying life stage specific habitat suitability criteria for depth, velocity, substrate and cover, the Physical Habitat Simulation system (PHABSIM) predicts depth and velocity across channel transects and combines these predictions with substrate or cover data into a habitat index known as weighted useable area (WUA) (Bovee, 1982; Milhous *et al.*, 1989). The WUA output is generally simulated for river reaches over a range of stream flows. Alternatively, two-dimensional (2-D) hydraulic and habitat models can be used to predict depth and velocity laterally and longitudinally throughout a length of river channel at a range of stream flows, and combine them with substrate or cover to predict the WUA for the site. Two-dimensional models have been suggested as a improvement and replacement for PHABSIM (Ghanem *et al.*, 1996; Leclerc *et al.*, 1995).

There are a number of potential advantages of using a 2-D model, versus PHABSIM. The use of a 2-D model avoids problems of where to place transects within a mesohabitat unit (Williams, 1996), since all of the mesohabitat unit is modeled with a 2-D model. Two-dimensional models have the potential to model depths and velocities in complex channels over a range of flows more accurately than PHABSIM because they take into account local bed topography and roughness, and explicitly use mechanistic processes (conservation of mass and momentum), rather than the reduced Manning's formulation and an empirical velocity adjustment factor (Leclerc *et al.*, 1995). Two-dimensional models can explicitly handle complex habitats, including transverse flows, across-channel variation in water surface elevations, and flow contractions/expansions, which cannot be modeled explicitly with PHABSIM (Ghanem *et al.*, 1996). Two-dimensional models can perform better than PHABSIM at representing patchy microhabitat features, such as gravel patches. The data can be collected with a stratified sampling scheme, with higher intensity sampling in areas with more complex or more quickly varying microhabitat features, and lower intensity sampling in areas with uniformly varying bed topography and uniform substrate. Bed topography and substrate mapping data can be collected at a very low flow, with the only data needed at high flow being discharge and water surface elevations at the top and bottom of the site and randomly sampled velocities for validation purposes.

In this paper, we evaluate whether the two-dimensional model used, River2D, (Steffler and Blackburn, 2001) is better than PHABSIM at predicting chinook salmon (*Oncorhynchus tschawytscha*) spawning habitat, and whether there are differences between PHABSIM and River2D in flow-habitat relationships for chinook salmon and steelhead (*Oncorhynchus mykiss*) spawning.

*This paper was prepared under the auspices of the U.S. Government and is therefore not subject to copywrite.
Received on March 22, 2007. Accepted on May 19, 2008.

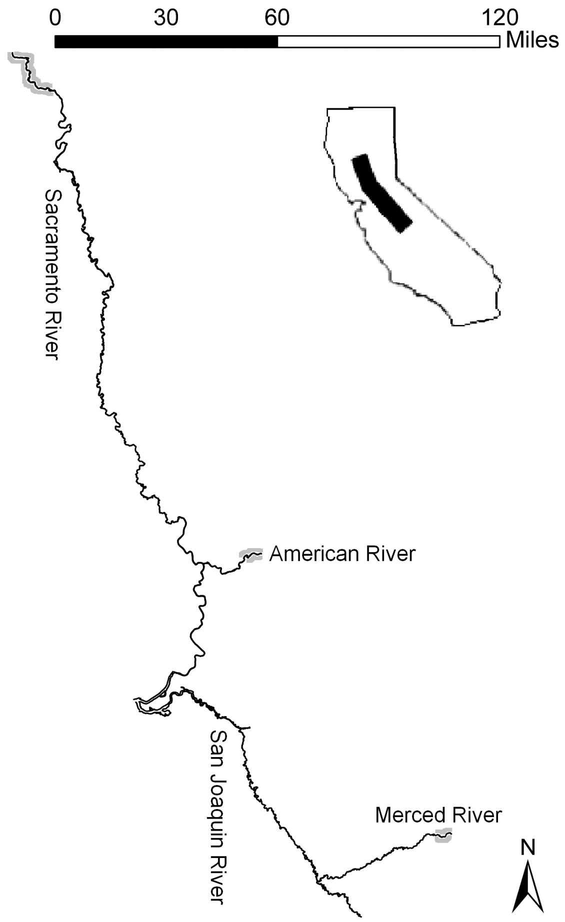


Figure 1 Location of the Sacramento, Merced and American Rivers, California. Shaded areas are the study reaches used to compare the spawning habitat predictions of the PHABSIM and River2D models.

2 Study sites

The Merced, American and Sacramento Rivers, located in the Central Valley of California, have a mean annual flow of 18.7, 106 and 275.8 m^3/s, respectively. This study was conducted in a 16-km reach of the Merced River, a 9-km reach of the American River, and a 47-km reach of the Sacramento River (Figure 1). PHABSIM and River2D were used to model one site on the Merced River, five sites on the American River and eight sites on the Sacramento River (Table 1). Three of the Sacramento River sites, located upstream of the Anderson-Cottonwood Irrigation District (ACID) Dam, were modeled for two conditions, with boards in or out at the ACID Dam. Stage at the sites was as much as 2 m higher with the boards in at the ACID, versus with the boards out.

3 Methods

3.1 Field measurements

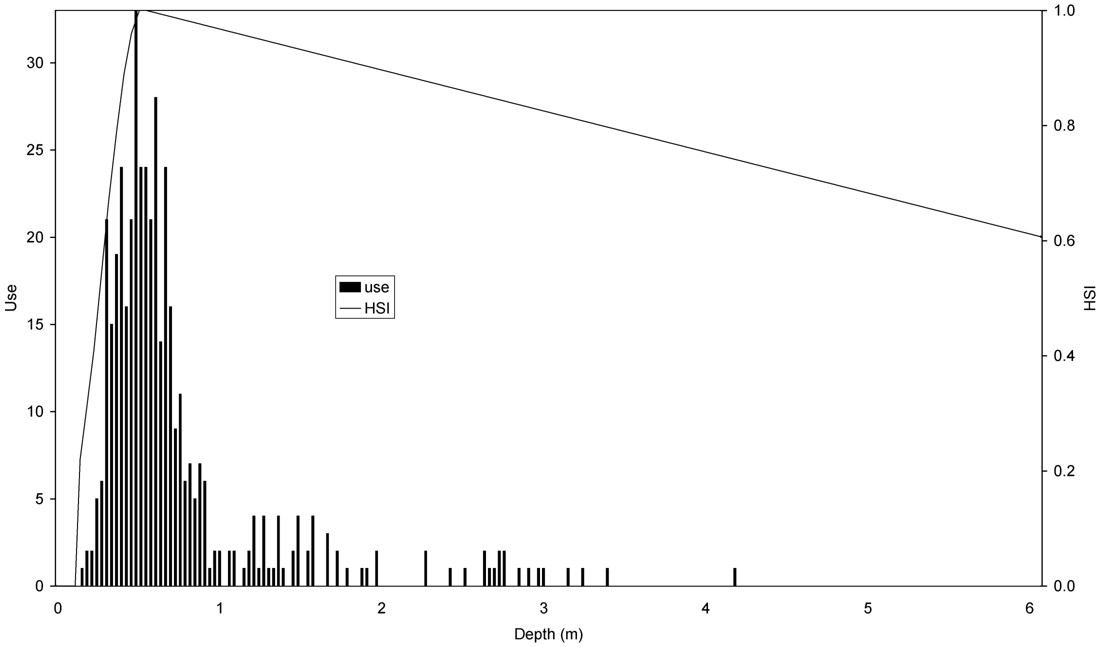
To model spawning habitat in the study sites, depth, velocity and substrate data were collected on 34 PHABSIM transects in the Sacramento River, 27 PHABSIM transects in the American River, and 6 PHABSIM transects in the Merced River, and substrate and bed topography data were collected for 2-dimensional hydraulic and habitat models for all 14 sites. For the PHABSIM transects, lateral cell boundaries were established systematically or where depth, velocity or substrate changed. Dominant substrate was visually assessed as the 2.5 to 5.0 cm size range of particles which comprised more than fifty percent of the surface area. For example, if more than fifty percent of the area was comprised of 5.0 to 10.0 cm particle sizes, the dominant substrate was classified as 5.0 to 10.0 cm. The midpoint of the dominant substrate size range would be an approximation of the D50 particle size. The substrate size classes used are shown in Figures 2 to 5. Depth, velocity and substrate data were collected in October 1996 at a flow of 11.95 m^3/s for the Merced River PHABSIM transects, in July to December 1998 at flows of 84.4 to 114.2 m^3/s for the American River PHABSIM transects, and in June to September 1997 at flows of 216.0 to 427.5 m^3/s for the Sacramento River PHABSIM transects. Water surface elevations and, for the Merced River, flows were measured at four to six flows for each PHABSIM transect. These flows ranged from 2.21 to 29.6 m^3/s for the Merced River during August to October 1996 (Gallagher and Gard, 1999), from 29.4 to 316.4 m^3/s for the American River during April to December 1998, and from 128.6 to 1192.5 m^3/s for the Sacramento River during May 1997 to March 1999 (Gard and Ballard, 2003). Flows for the American and Sacramento Rivers were determined from gage readings.

Table 1 Characteristics of study sites. Three of the Sacramento River sites were modeled for two conditions – with boards in and out at the Anderson-Cottonwood Irrigation District (ACID) Dam. Stage at the study sites was up to 2 m higher with the ACID Dam boards in, versus with the boards out. The Merced site was simulated for 11 flows, one of the American River sites (El Manto) was simulated for 35 flows, and the Sacramento sites and the rest of the American River sites were simulated for 30 flows. The lower end of the simulated flow range for the El Manto site was 14.2 m^3/s.

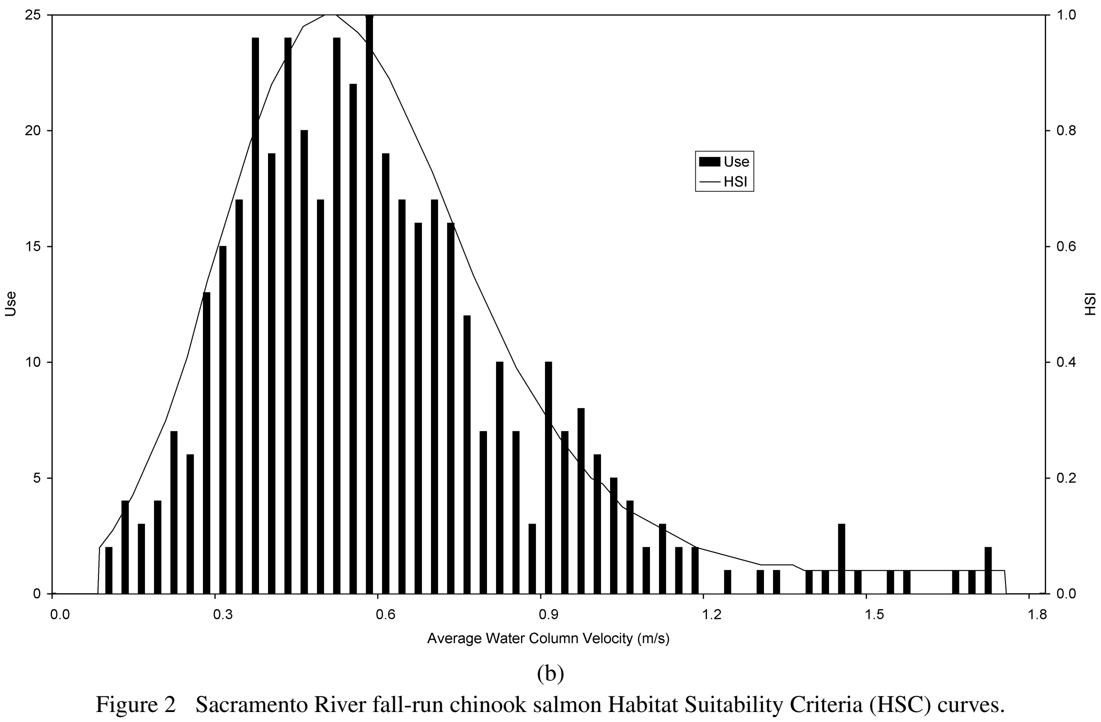
River	Number of sites	Number of transects/ site	Length of site (channel widths)	Range of simulated flows (m^3/s)
Sacramento	8	1–10	0.33–1.88	92.0–877.8
American	5	2–7	2.43–10.43	28.3–311.5
Merced	1	6	2.03	5.7–19.8

The downstream-most and upstream-most PHABSIM transects were used for, respectively, the bottom and top of each River2D site. The remaining PHABSIM transects were used to establish a portion of the bed topography and substrate distribution of each River2D site. Data to develop the rest of the bed topography and substrate distribution of the River2D sites were collected with a total station for all of the Merced River site and the dry and shallow portions of the American and Sacramento River sites, generally in sets of points going across the channel. Data for the bed topography and substrate distribution of the deep (greater than 1 m depth) portions of the American and Sacramento River sites were collected with an Acoustic Doppler Current Profiler (ADCP) and underwater video (Gard and Ballard, 2003). The average density of points from all sources (PHABSIM transects, ADCP and total station) used to develop the bed topography for the River2D model was 2.65 points/100 m^2 (Table 2). The stage-discharge relationship for the downstream-most PHABSIM transect and the flows at the upstream boundary were used as inputs to the River2D model of each site, while the water surface elevation measured at the highest flow at the remaining PHABSIM transects were used to calibrate the River2D model of each site by adjusting the bed roughnesses of the site until the water surface elevations predicted by River2D matched the measured water surface elevations.

To develop chinook salmon spawning habitat suitability criteria, depth, velocity and substrate data were collected on fall-run chinook salmon redds in the Merced, American and Sacramento Rivers and on late-fall-run and winter-run chinook salmon redds in the Sacramento River (Table 3). The methods used to collect habitat suitability criteria for the Merced and American Rivers are given in Gard (1998), while the methods used to collect habitat suitability criteria for the Sacramento River are given in Gard and Ballard (2003). Horizontal surveying was used to determine the location of redds in the Merced River site in 1996 and in two of the American River sites on December 14–17, 1998, and a Global Positioning System (GPS) receiver was used to determine the location of redds in all of the Sacramento River sites (occupied n values in Tables 4 and 5).

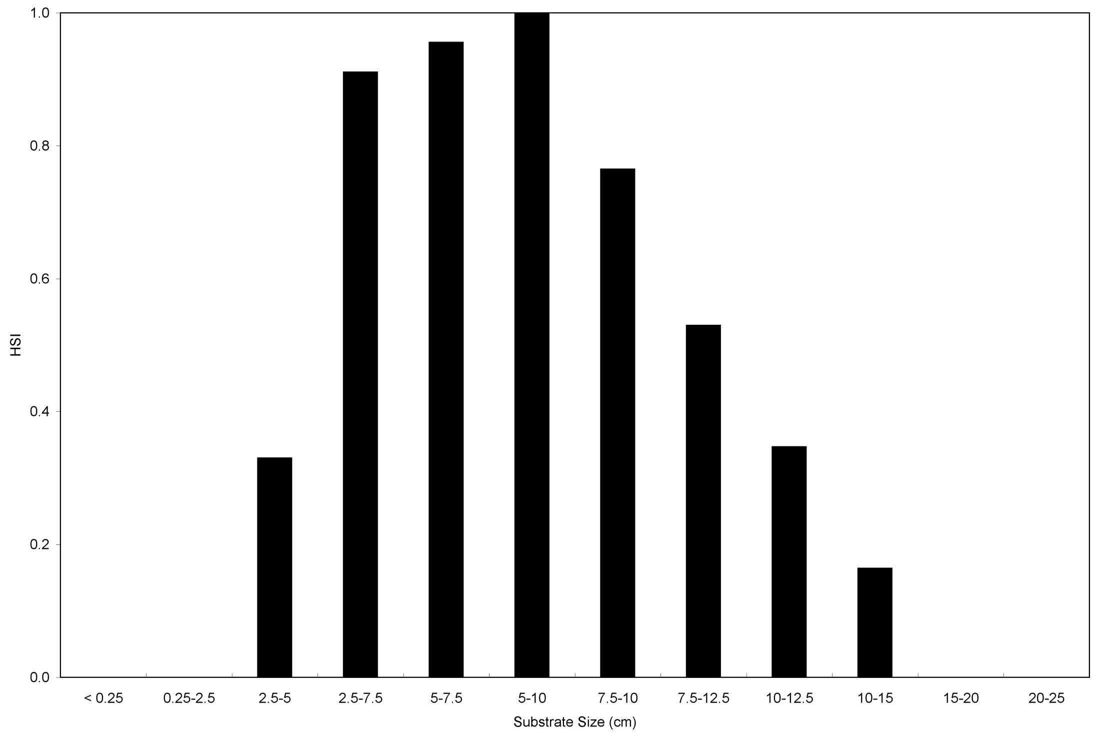


(a)



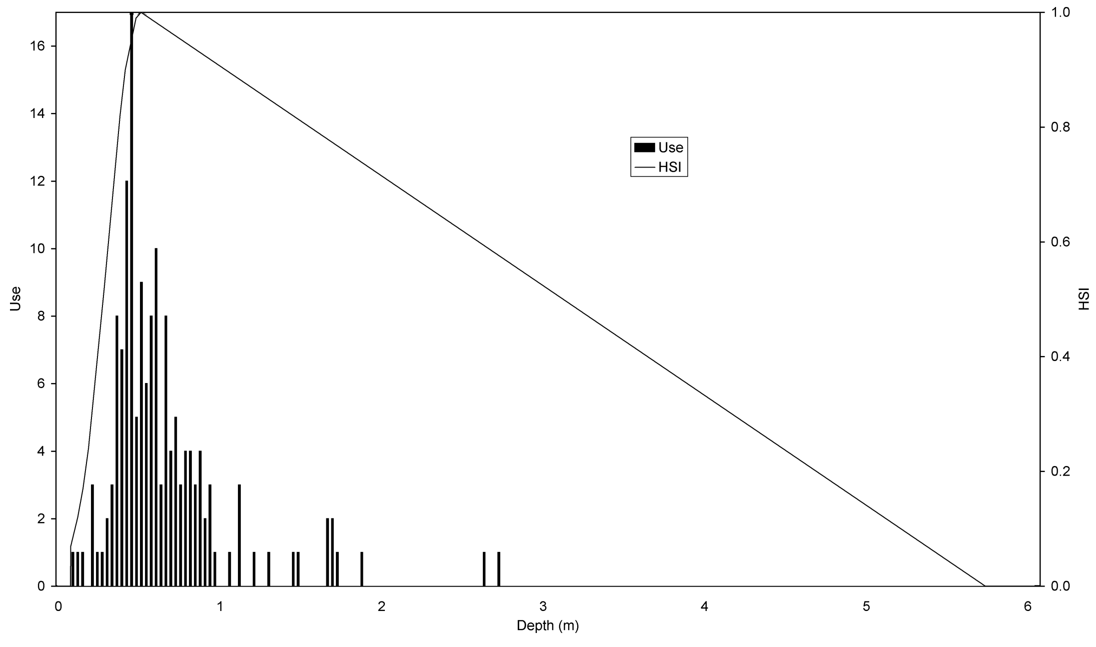
(b)

Figure 2 Sacramento River fall-run chinook salmon Habitat Suitability Criteria (HSC) curves.



(c)

Figure 2 (*Continued*)



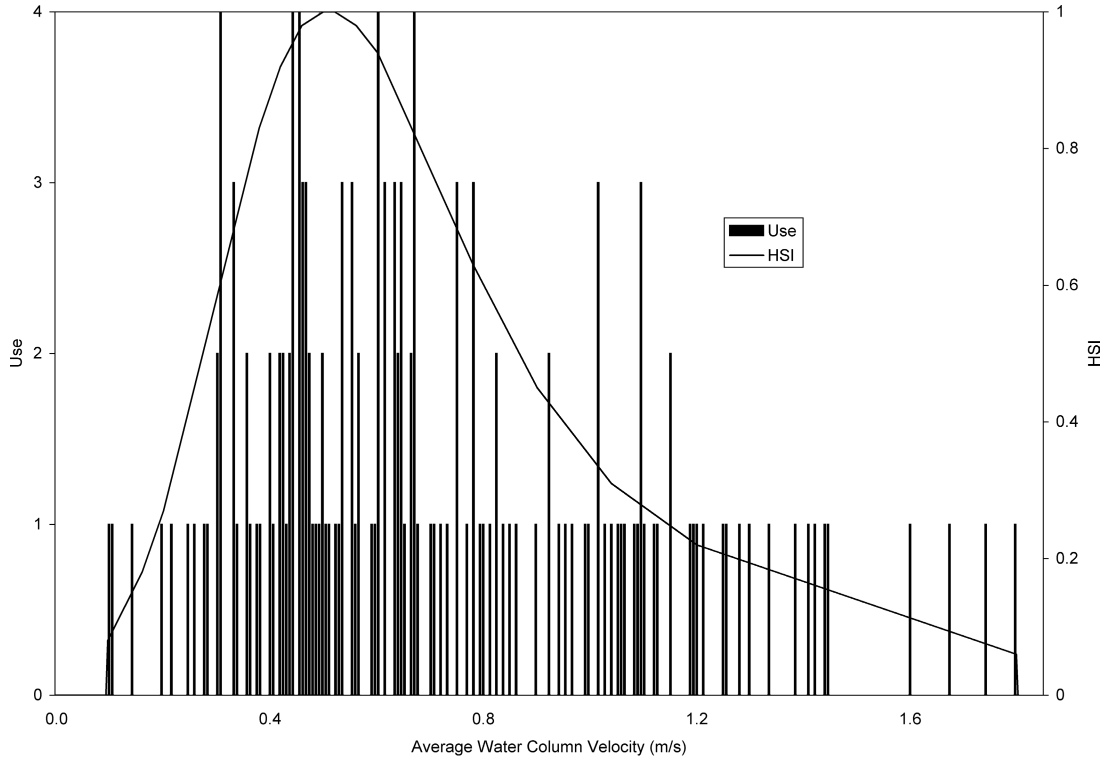
(a)

Figure 3 Sacramento River late-fall-run chinook salmon Habitat Suitability Criteria (HSC) curves.

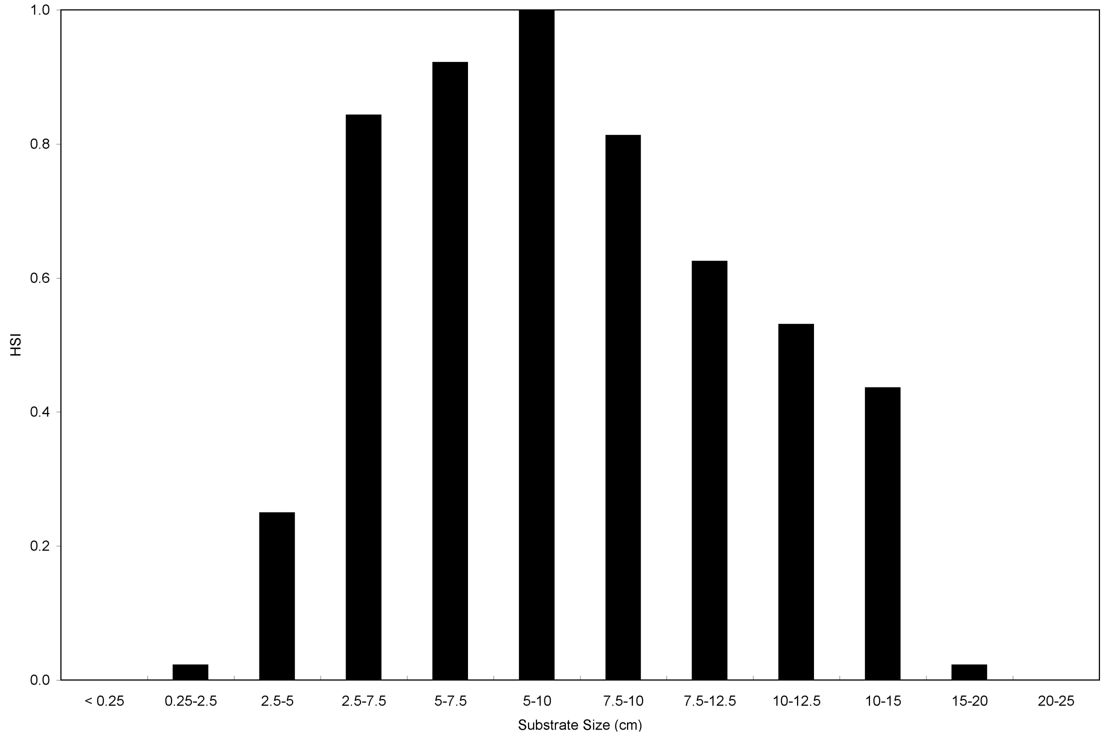
3.2 Habitat modeling

Average water column velocities, water surface elevations, riverbed elevations, cell substrate categories, and site discharges were entered into PHABSIM to create hydraulic models for each transect. PHABSIM hydraulic data were calibrated following procedures in Milhous *et al.* (1989). These procedures involve the development of stage-discharge relationships using three possible techniques: a log-log linear rating curve, Manning's equation, or a step-backwater method. The calibrated files for each site were used in PHABSIM to simulate hydraulic characteristics for the range of flows in Table 1, and for the average flows each year from the beginning of spawning through the end of redd data collection (Table 6).

The River2D model solves the two-dimensional, depth averaged St. Venant equations expressed in conservative form (Steffler and Blackburn, 2002). The River2D model uses a finite



(b)



(c)

Figure 3 (*Continued*)

element numerical method based on the Streamline Upwind Petrov-Galerkin weighted residual formulation, using a Newton Raphson iterative method (Steffler and Blackburn, 2002). The River2D model achieves turbulence closure through the use of a Boussinesq type eddy viscosity formulation (Steffler and Blackburn, 2002). The basis for the current form of RIVER2D is given in Ghanem *et al.* (1995).

Bed topography, bed roughness and substrate distribution data were entered into River2D to create hydraulic models for each site. To minimize the effects of inflow boundary condition specifications, a one-channel-width upstream artificial extension was added to each site by translating the cross-sectional topography at the top of the site upstream parallel to the top PHABSIM transect, with a bedslope equal to the water surface elevation slope

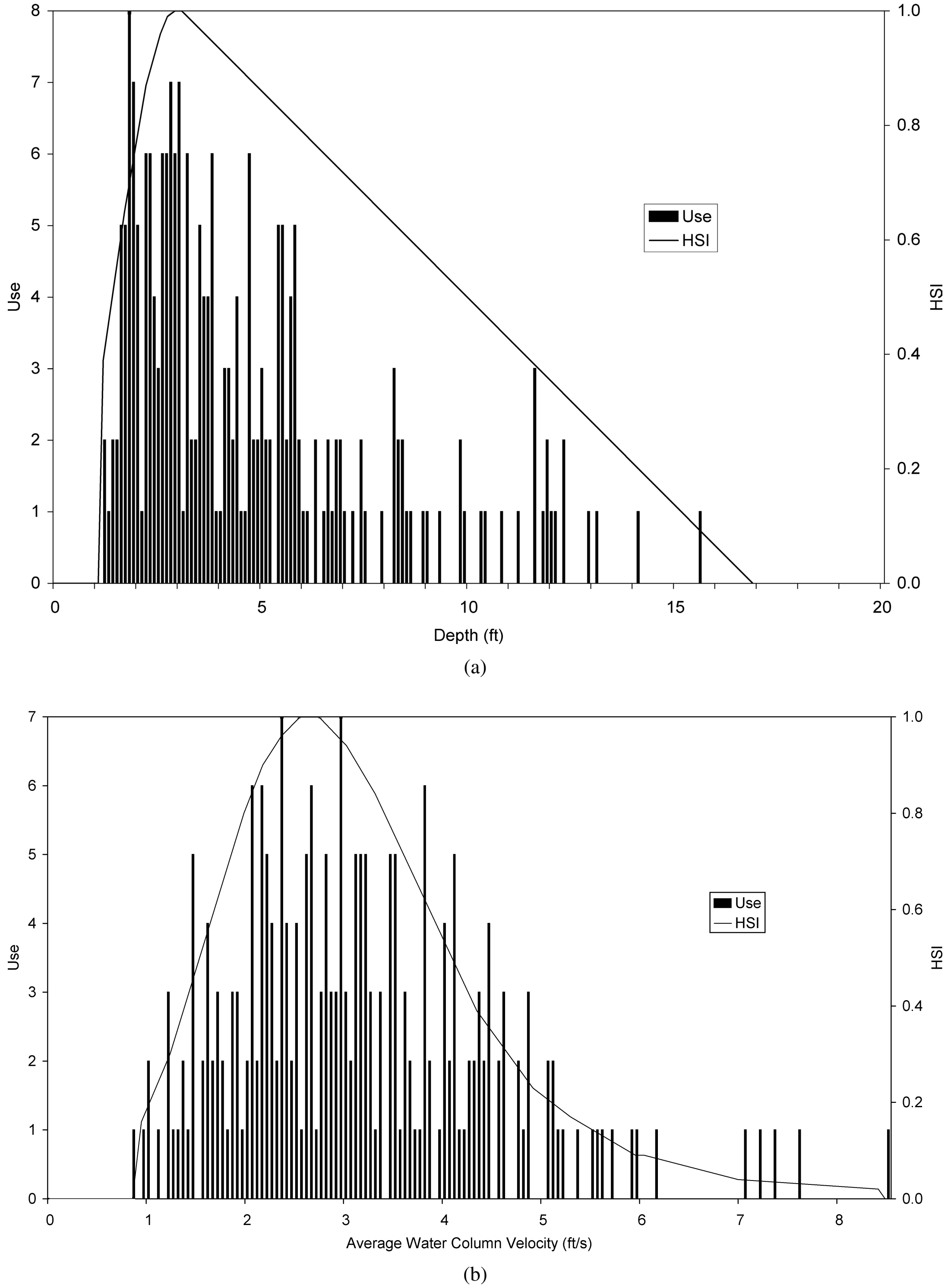
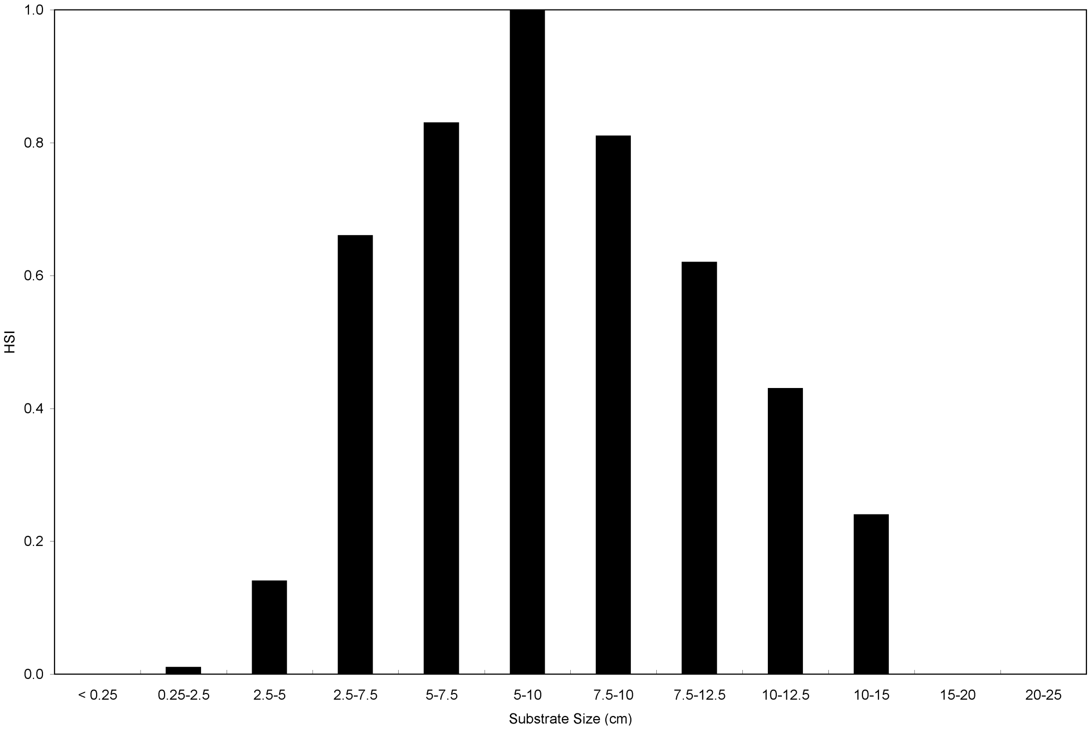


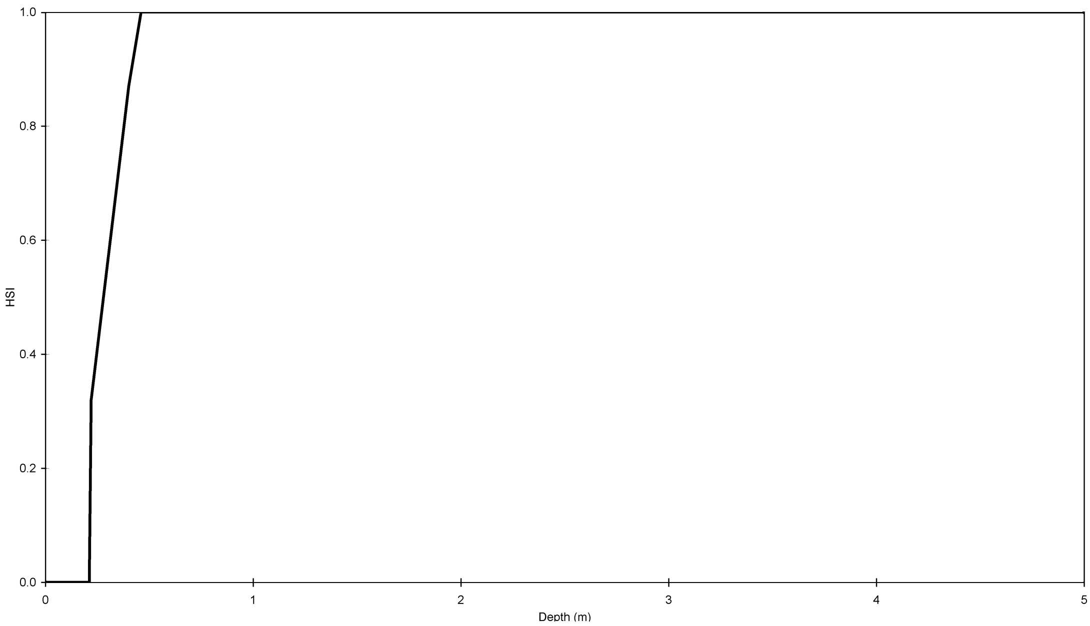
Figure 4 Sacramento River winter-run chinook salmon Habitat Suitability Criteria (HSC) curves.

at the top of the site. The River2D model distributes flow across the inflow boundary proportional to depth, resulting in the fastest velocity being at the thalweg. The River2D model used a triangular irregular network (TIN) grid, with grid elements ranging in size from 13 m in areas with uniform topography to 0.7 m in areas with rapidly varying topography (Figure 6). The grid element sizes were selected to minimize the elevation error between the TIN and the underlying bed topography data, while taking into account computational limitations of large numbers of grid elements. The number of grid elements, from site to site, ranged from 5,475 to 24,488. River2D hydraulic data were calibrated by adjusting bed roughnesses until simulated water surface elevations matched measured water surface elevations. The initial values of bed roughness for the River2D model were set equal to five times the midpoint of the substrate range, i.e. a substrate range of 5 to 10 cm would have an initial bed roughness of 0.4 m (7.5 cm $\times$ 5). Five times the average particle size is approximately the same as 2 to 3 times the d85 particle size, which is recommended as an estimate of bed roughness height (Yalin 1977). The bed roughnesses were adjusted by applying a fixed multiplier



(c)

Figure 4 (*Continued*)

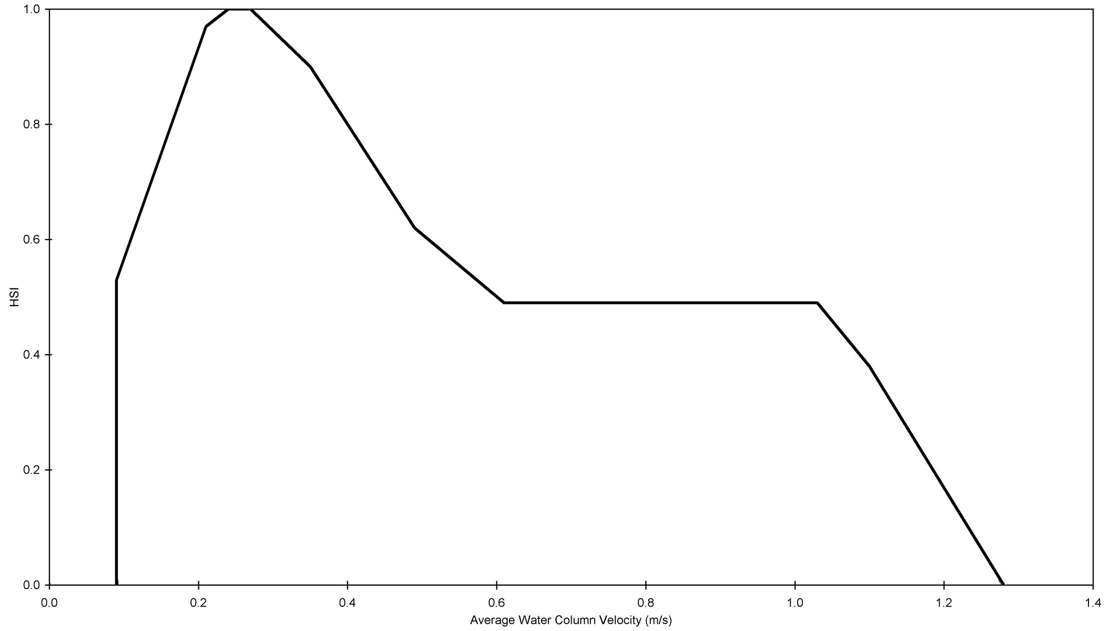


(a)

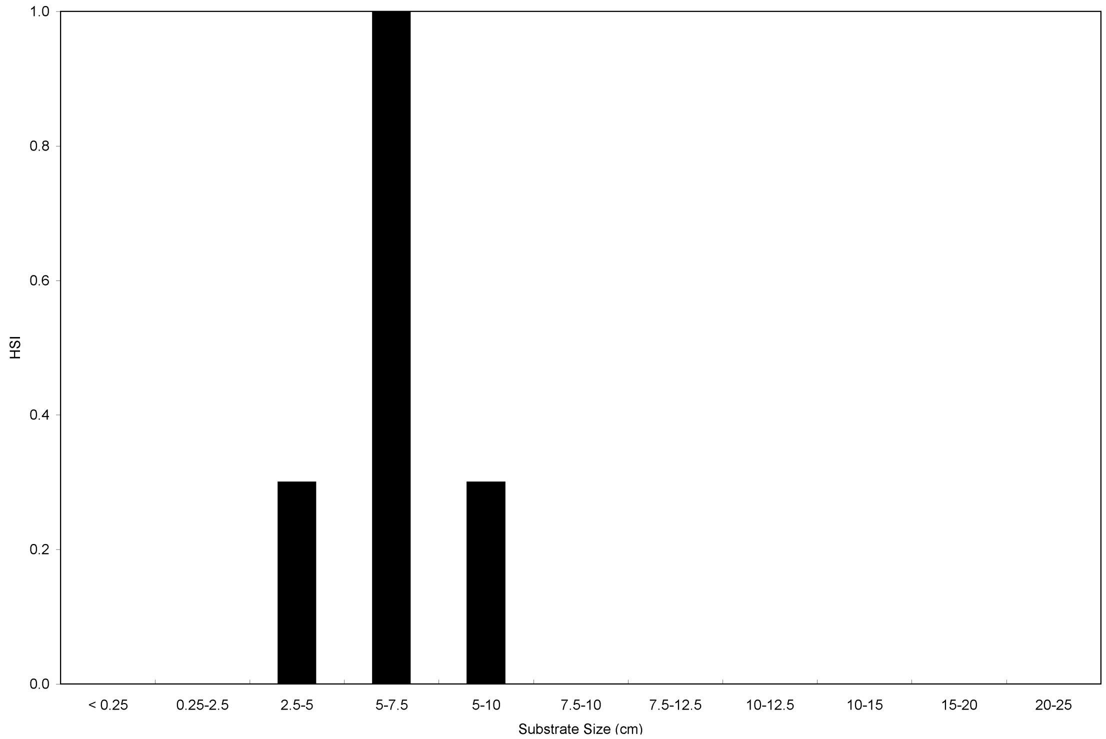
Figure 5 Steelhead Habitat Suitability Criteria (HSC) curves used to simulate steelhead spawning habitat in the Sacramento and Lower American Rivers.

to all of the bed roughnesses. The values of all other River2D hydraulic parameters were left at their default values (upwinding coefficient = 0.5, minimum groundwater depth = 0.1 m, groundwater transmissivity = 0.1 m^2/s, groundwater storativity = 1, and eddy viscosity parameters epsilon1 = 0.01 m^2/s, epsilon2 = 0.5 m^2/s and epsilon3 = 0.1 m^2/s). The upwinding coefficient is used in River2D's Petrov-Galerkin finite element scheme, the groundwater parameters are used for River2D's wetting/drying algorithm, and the eddy viscosity parameters are used in River2D's transverse shear model (Steffler and Blackburn, 2002). The calibrated files for each site were used in River2D to simulate hydraulic characteristics for the range of flows in Table 1, and for the average flows each year from the beginning of spawning through to the end of redd data collection (Table 6).

Habitat suitability curves (HSC) are used in PHABSIM and River2D to translate hydraulic and structural elements of rivers into indices of habitat quality called combined suitability indices (CSI), calculated as the product of the depth, velocity and



(b)



(c)

Figure 5 (*Continued*)

Table 2 Study site data collection. There is only one value for the range of point densities for the Merced River since there was only one study site on that river.

River	Range of point densities (points/100 m^2)	Number of points per reach
Sacramento	0.90–4.16	4717
American	1.03–1.24	4784
Merced	3.41	367

substrate suitabilities. The habitat suitability criteria data for the Merced and Lower American Rivers in Table 3 were used to develop HSC for fall-run chinook salmon in the Merced and Lower American Rivers (Gard, 1998). The habitat suitability criteria data in Table 3 for the Sacramento River were used to develop HSC for fall-run, late-fall-run and winter-run chinook salmon in the Sacramento River (Figures 2 to 4) using the techniques in Gard (1998). Habitat suitability criteria for steelhead (Figure 5) were developed from depth and velocity

Table 3 Habitat suitability criteria data collected as part of this study. Flows are the range of flows during data collection. Spawning criteria for late-fall chinook salmon were developed using the data from this study and data collected on 79 redds by the California Department of Fish and Game on Jan 1–Mar 3 1986–1988 at flows of 89.2 to 162.8 m^3/s.

River	Race	Number of Redds	Data collection dates	Flow (m^3/s)
Sacramento	Fall-run	437	Oct 23–Nov 25 1995–1999	130.4–176.8
Sacramento	Late-fall-run	77	Feb 27–Mar 29 2001	90.2–117.0
Sacramento	Winter-run	227	May 26–Jul 15 1996–2001	297.2–563.8
American	Fall-run	218	Nov 6–7 1996	78.6
Merced	Fall-run	186	Nov 12–14 1996	7.79

data collected on steelhead redds in the Lower American River by the California Department of Fish and Game and substrate data collected on steelhead redds in the Trinity River by the U.S. Fish and Wildlife Service using the methods in Gard (1998).

The calibrated PHABSIM and River2D hydraulic simulations were used with the above HSC to generate flow-habitat relationships for fall-run chinook salmon spawning in the Sacramento, American and Merced River sites, for steelhead spawning in the Sacramento and American River sites, and for late-fall-run and winter-run chinook salmon spawning in the Sacramento River sites. The calibrated PHABSIM hydraulic simulations for the flows in Table 6 were used with the chinook salmon HSC to calculate the CSI values predicted by PHABSIM for occupied (cells with redds) and unoccupied cells for each site and year where redd locations were determined. For unoccupied cells, all wetted cells

Table 4 Results of Mann-Whitney U Tests for PHABSIM occupied versus unoccupied cells.

River	Race	Occupied n	Unoccupied n	Occupied median	Unoccupied median	p-value
Merced	Fall	28	221	0.10	0.00	0.011
American	Fall	103	497	0.23	0.01	0.003
Sacramento	Fall	71	3081	0.31	0.01	< 0.000001
Sacramento	Late-fall	22	1906	0.26	0.17	0.16
Sacramento	Winter	51	6164	0.29	0.00	< 0.000001

Table 5 Results of Mann-Whitney U Tests for 2-D model occupied versus unoccupied locations.

River	Race	Occupied n	Unoccupied n	Occupied median	Unoccupied median	p-value
Merced	Fall	33	220	0.54	0.27	0.001
American	Fall	184	458	0.04	0.00	0.000003
Sacramento	Fall	74	3080	0.11	0.03	0.000026
Sacramento	Late-fall	16	1906	0.07	0.14	0.313
Sacramento	Winter	58	6164	0.14	0.01	0.000062

Table 6 Time period and average chinook salmon spawning river discharge (m^3/s) for the Merced, Lower American and Sacramento Rivers. Data are only given for years in which redd locations were recorded for study sites. The range of flows for the Sacramento River sites reflects the different flows present at different sites due to tributary inflow within the reach and differences from site to site in the final date of redd data collection.

	Race	1996	1997	1998	1999	2000	2001
Merced	Fall						
Time period		10/23–11/14					
Average		8.4					
American	Fall						
Time period				11/11–12/17			
Average				87.4			
Sacramento	Fall						
Time period			10/9–11/20			10/7–11/4	
Average			127.9–130.3			173.3–177.8	
Sacramento	Late-						
Time period	fall						1/6–3/29
Average							108.1–117.0
Sacramento	Winter						
Time period				5/15–6/23	4/15–7/14	4/15–7/10	4/15–6/21
Average				445.4–469.4	288.6–308.5	308.1–324.0	281.2

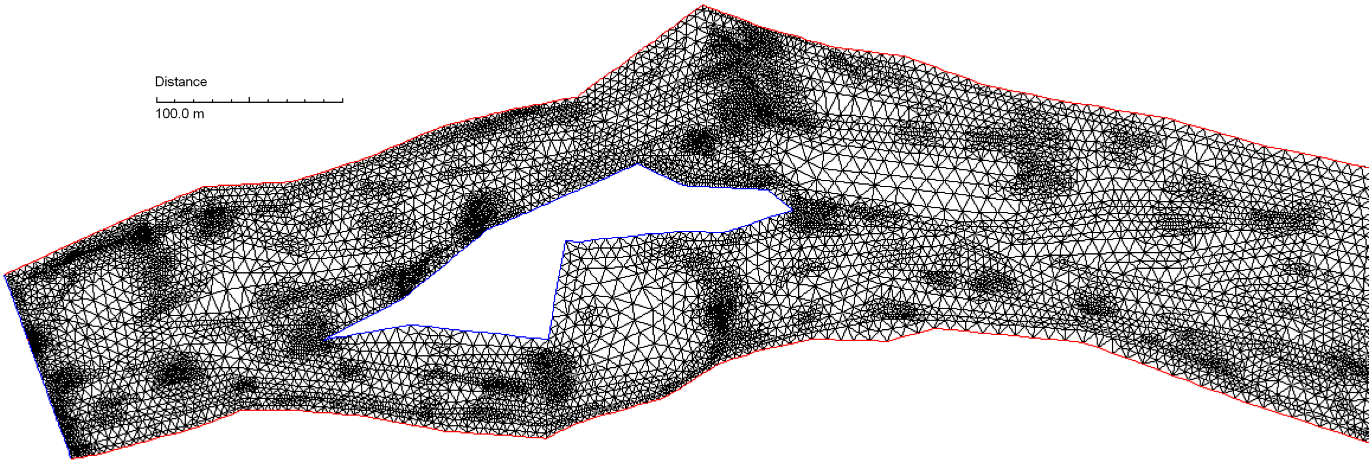


Figure 6 Example of triangular irregular element mesh used to perform the two-dimensional hydraulic modeling of the American River.

were used. Similarly, the calibrated River2D simulations for the flows in Table 6 were used with the same chinook salmon HSC to calculate the CSI values predicted by River2D for occupied and unoccupied locations for each site and year where redd locations were determined. Unoccupied locations were randomly selected which met the following criteria: they were farther than one m from an occupied location, and they were wetted. The number of unoccupied River2D locations (Table 5) was chosen to be similar to the number of unoccupied PHABSIM cells (Table 4). The number of occupied River2D locations (Table 5) differs from the number of occupied PHABSIM cells (Table 4) for the following reasons: 1) some PHABSIM cells contained more than one redd, while each occupied River2D location only contained one redd; 2) some portions of the River2D sites were not represented by any of the PHABSIM transects; and 3) redds located upstream of the uppermost PHABSIM transect, but within the portion of the channel represented by the uppermost PHABSIM transect, would be located within PHABSIM cells but would be upstream of the River2D site. Model type (River2D versus PHABSIM) came into the analysis of CSI because the analysis used the CSI calculated by the two models based on the depths, velocities and substrates predicted by each model at the redd locations, rather than the CSI that could be calculated from the measured depths, velocities and substrates. The River2D model calculates CSI using the depths and velocities from the hydraulic simulation, substrate data from a channel index file, and the HSC. The key differences between the models tested in this paper are that PHABSIM is a one-dimensional model that simulates velocities using Manning's n values, while River2D is a two-dimensional model that simulates velocities using conservation of mass and momentum. During the habitat calculations, substrate is assigned to each River2D node based on the nearest substrate datapoint in the channel index file (either longitudinally or laterally), while PHABSIM, with longitudinal cells, assigns substrate values based on the nearest vertical longitudinally.

3.3 Data analysis

Mann-Whitney U tests (Wilkinson, 1990) were used to determine for each river, and, in the case of the Sacramento River, for each race of chinook salmon, if there was a significant difference in the CSI predicted by PHABSIM for occupied versus unoccupied cells, and if there was a significant difference in the CSI predicted by River2D for occupied versus unoccupied locations. This test is analagous to the transferability test described by Thomas and Bovee (1993). Kolmogorov-Smirnov tests (Steel and Torrie, 1980) were performed for each site for each set of suitability criteria to detemine if there was a significant difference between the PHABSIM and River2D flow-habitat relationships. Separate Kolmogorov-Smirnov tests were performed for the three Sacramento River sites upstream of the ACID dam for the two conditions simulated (boards in or out at the ACID Dam). As a result, there were a total of 55 Kolmogorov-Smirnov tests ([3 Sacramento River sites above ACID Dam $\times$ 2 conditions + 5 Sacramento River sites below ACID Dam] $\times$ 4 HSC sets + 5 American River sites $\times$ 2 HSC sets + 1 Merced River site $\times$ 1 HSC set).

4 Results

Velocity validation statistics of the River2D hydraulic model are given in Table 7, while a graphical example of the validation results are shown in Figure 7. Typical results of the River2D habitat model are shown in Figure 8. The CSI of occupied locations predicted by both PHABSIM (Table 4) and River2D (Table 5) was significantly greater than the CSI of unoccupied locations at $p = 0.05$ (Mann-Whitney U test) for fall-run chinook salmon spawning for all three rivers and for winter-run chinook salmon spawning in the Sacramento River. However, the CSI of occupied locations predicted by both PHABSIM and River2D were not significantly different from the CSI of unoccupied locations at $p = 0.05$ (Mann-Whitney U test) for late-fall-run chinook salmon spawning in the Sacramento River. The number of occupied cells and locations for late-fall-run (Tables 4 and 5) was lower than for the other Mann-Whitney U tests. The median CSI predicted for redd locations by River2D was greater than that predicted by PHABSIM for the Merced River, but was less for the American and Sacramento Rivers (Tables 4 and 5). The percentage of occupied locations where River2D predicted a CSI of 0 was less than

Table 7 River2D hydraulic modeling validation results. The errors were calculated as the absolute value of the difference between the measured and simulated velocities.

River	Site number	Mean error (m/s) for velocities < 0.91 m/s	Mean error (%) for velocities > 0.91 m/s
Sacramento	1	0.31	24%
Sacramento	2	0.17	17%
Sacramento	3	0.14	16%
Sacramento	4	0.52	30%
Sacramento	5	0.29	15%
Sacramento	6	0.22	13%
Sacramento	7	0.48	13%
Sacramento	8	0.34	20%
American	1	0.63	38%
American	2	0.25	27%
American	3	0.27	17%
American	4	0.35	24%
American	5	0.31	22%
Merced	1	0.17	26%

the percentage of occupied cells where PHABSIM predicted a CSI of 0 for fall-run chinook salmon spawning in all three rivers, but was greater for late-fall-run and winter-run chinook salmon spawning (Tables 7 and 8). For both PHABSIM and River2D, a substrate which was too large or small was the cause of most of the occupied locations which were predicted to have a CSI of 0 (Tables 7 and 8).

The Kolomogorov-Smirnov D statistics for the comparisons of PHABSIM and River2D flowhabitat relationships (Figure 9) ranged from 0.007 (Figure 10C) to 0.41 (Figure 10A), with a median value of 0.07 (Figure 10B). Only one PHABSIM flow-habitat relationship (Figure 10A) was significantly different from the River2D flow habitat relationship at $p = 0.05$. Even though the differences between the PHABSIM and River2D flow habitat relationships were almost allways not statistically significantly different, differences in the flow habitat relationships between the two model could result in different flow management decisions. For example, a comparison with a relatively low Kolomogorov-Smirnov D statistic of 0.03 (Figure 10D) has a maximum amount of spawning habitat at 85.0 m^3/s for PHABSIM, versus at 118.9 m^3/s with River2D, a 40 percent higher flow.

5 Discussion

Errors in the habitat predictions for occupied locations for PHABSIM can be due to longitudinal variation in depth, velocity and substrate (Gallagher and Gard, 1999) or due to the velocity distribution across the channel changing with flow. Errors in the habitat predictions for occupied locations for River2D can be due to inadequate detail in mapping substrate distribution, insufficient data collected to correctly map the bed topography of the site, or effects of the bed topography upstream of the study site not being included in the model. For the Sacramento River sites, a substantial proportion of the error for both the PHABSIM and

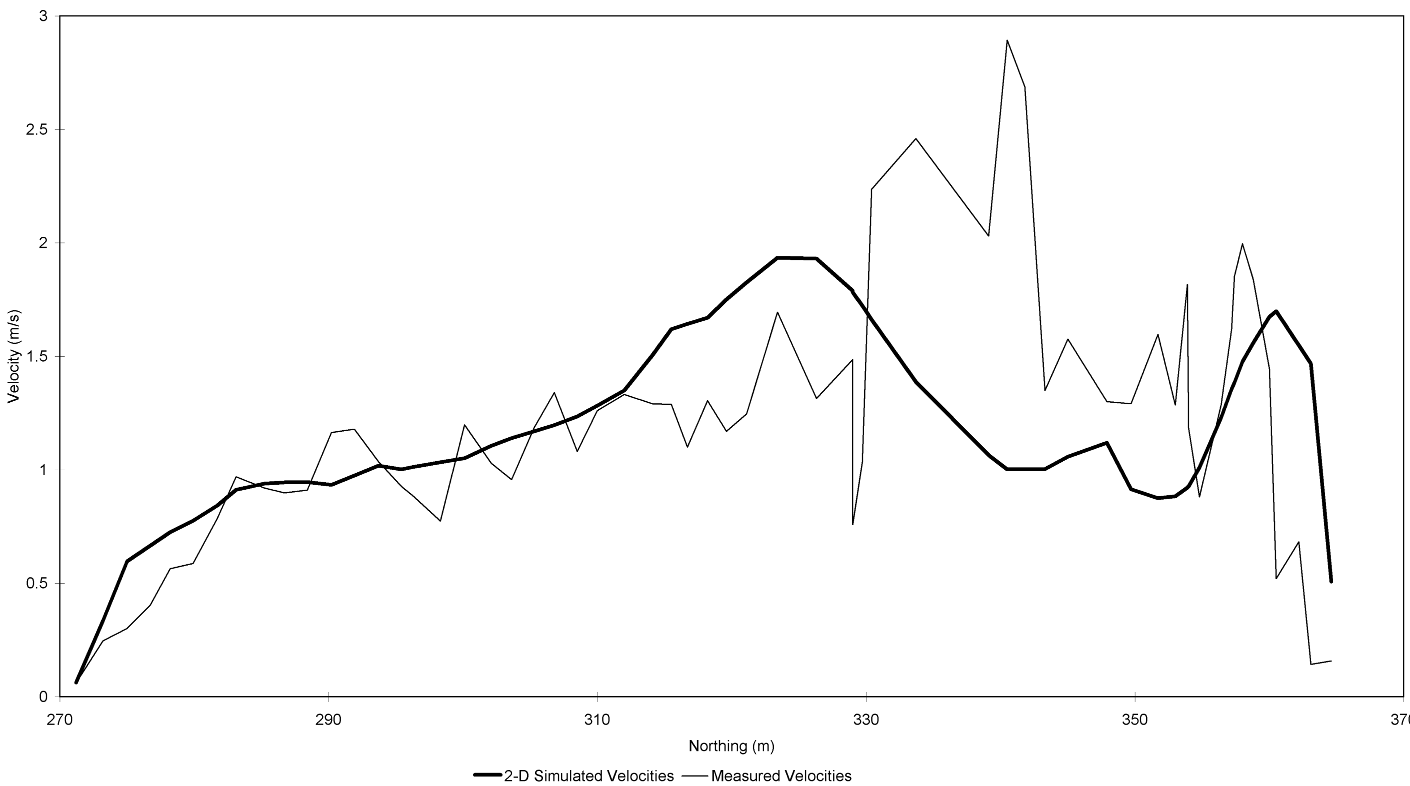


Figure 7 Example of River2D validation for one of the transects of the American River site illustrated in Figure 6 at a flow of 88.2 m^3/s.

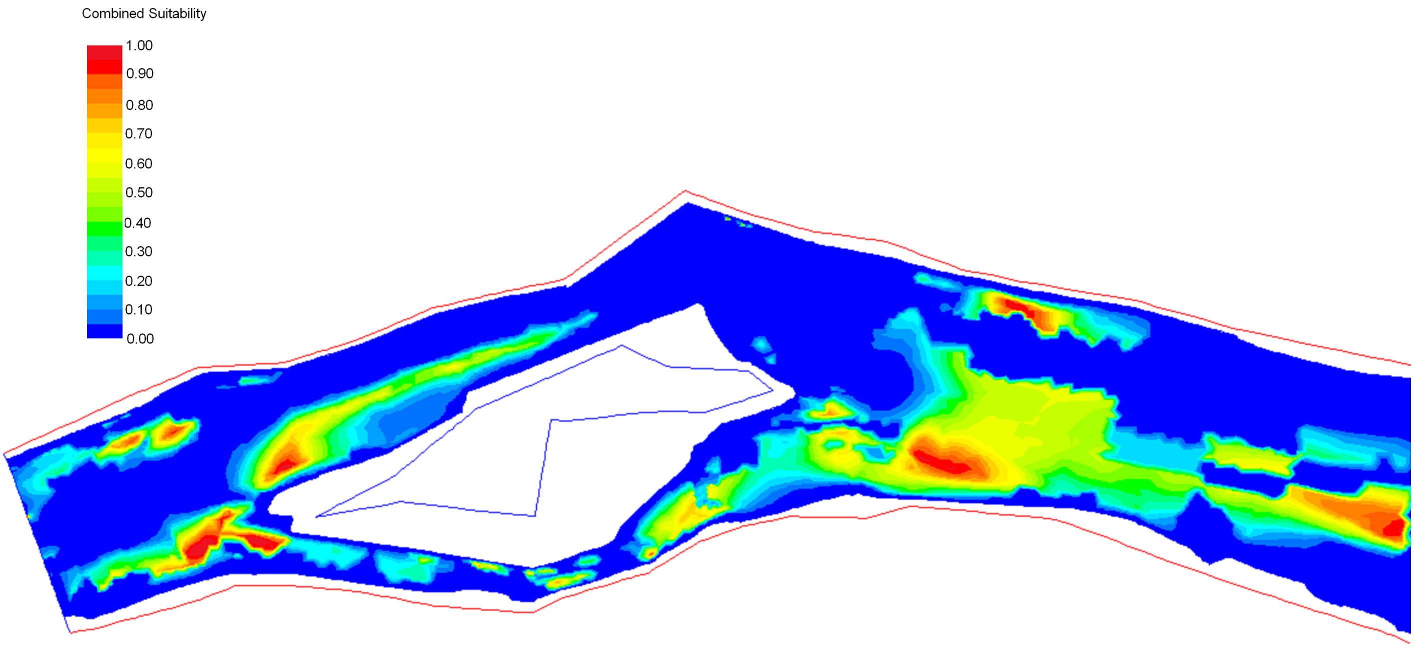


Figure 8 Example of River2D output of CSI for fall-run chinook salmon spawning at a flow of 87.8 m^3/s for the American River site illustrated in Figure 6.

Table 8 Characteristics of occupied cells predicted by PHABSIM. The numbers in the last five columns are the number of occupied cells that PHABSIM predicted having a CSI of 0 as a result of the cause given for that column. The percent of occupied cells with a CSI of 0 is the total number of occupied cells with a CSI of 0 (incuding all of the causes in the last five columns) divided by the total number of occupied cells (as given in Table 4).

River	Race	% Occupied cells with CSI of 0	Substrate too large or small	Dry	Too shallow	Too slow	Too fast
Merced	Fall	4%	1	0	0	0	0
American	Fall	36%	24	7	1	0	5
Sacramento	Fall	28%	16	4	0	0	0
Sacramento	Late-Fall	18%	3	1	0	0	0
Sacramento	Winter	22%	11	0	0	0	0

River2D models habitat predictions can be attributed to errors in the GPS measurements of redd locations, rather than errors in the habitat predictions of the models. The location of redds indicated by the GPS measurement can be as much as 5 m from the actual redd location (Gard and Ballard, 2003). In several cases, the redd location indicated by the GPS measurement was up onto the riverbank above water's edge.

The ability of PHABSIM in this case to relatively accurately predict the CSI of redd locations can be attributed to the number and spacing of transects, such that conditions at the transect tended to be representative of the depths, velocities and substrates present throughout the cells, and because flow at the sites chosen is largely one-dimensional, with only limited two-dimensional effects, such as transverse flows and across-channel variation in water surface elevations. There is a balance in the predictive accuracy of PHABSIM and River2D between the shapes of cells and the velocity information provided to each model. River2D will tend to be more accurate than PHABSIM because of the smaller triangular elements used by River2D, compared to the large rectagular cells used by PHABSIM. At least at flows close to those at which velocity data were collected and at locations close to the transect, PHABSIM will typically do a good job in predicting velocities, since it calculates the Manning's n value for each cell from the measured depth and velocity, and then calculates the simulated velocity from the Manning's n value. In contrast, River2D does not use any measured velocity data to predict velocities. While the only way to improve the performance of the PHABSIM habitat predictions would have been to increase the number of transects, and thus decrease the longitudinal length of the cells, there are several techniques that could have been used to improve the performance of the River2D habitat predictions with the existing dataset. It appears based on our substrate data that substrate varies more laterally than longitudinally. To test whether this supposition could be used to improve the performance of River2D, a test channel index file was created for the American River site in Figures 6 and 8 in which longitudinal breaklines were added to force River2D to predict substrate at a given location based on the nearest longitudinal point where substrate data was collected. This decreased the number of redds with predicted substrate suitability of zero

Table 9 Characteristics of occupied locations predicted by River2D. The numbers in the last five columns are the number of occupied locations that River2D predicted having a CSI of 0 as a result of the cause given for that column. The percent of occupied locations with a CSI of 0 is the total number of occupied locations with a CSI of 0 (incuding all of the causes in the last five columns) divided by the total number of occupied locations (as given in Table 5).

River	Race	% Occupied cells with CSI of 0	Substrate too large or small	Dry	Too shallow	Too slow	Too fast
Merced	Fall	0%	0	0	0	0	0
American	Fall	33%	52	5	0	1	3
Sacramento	Fall	22%	13	1	1	0	1
Sacramento	Late-fall	37%	6	0	0	0	0
Sacramento	Winter	34%	13	0	4	3	0

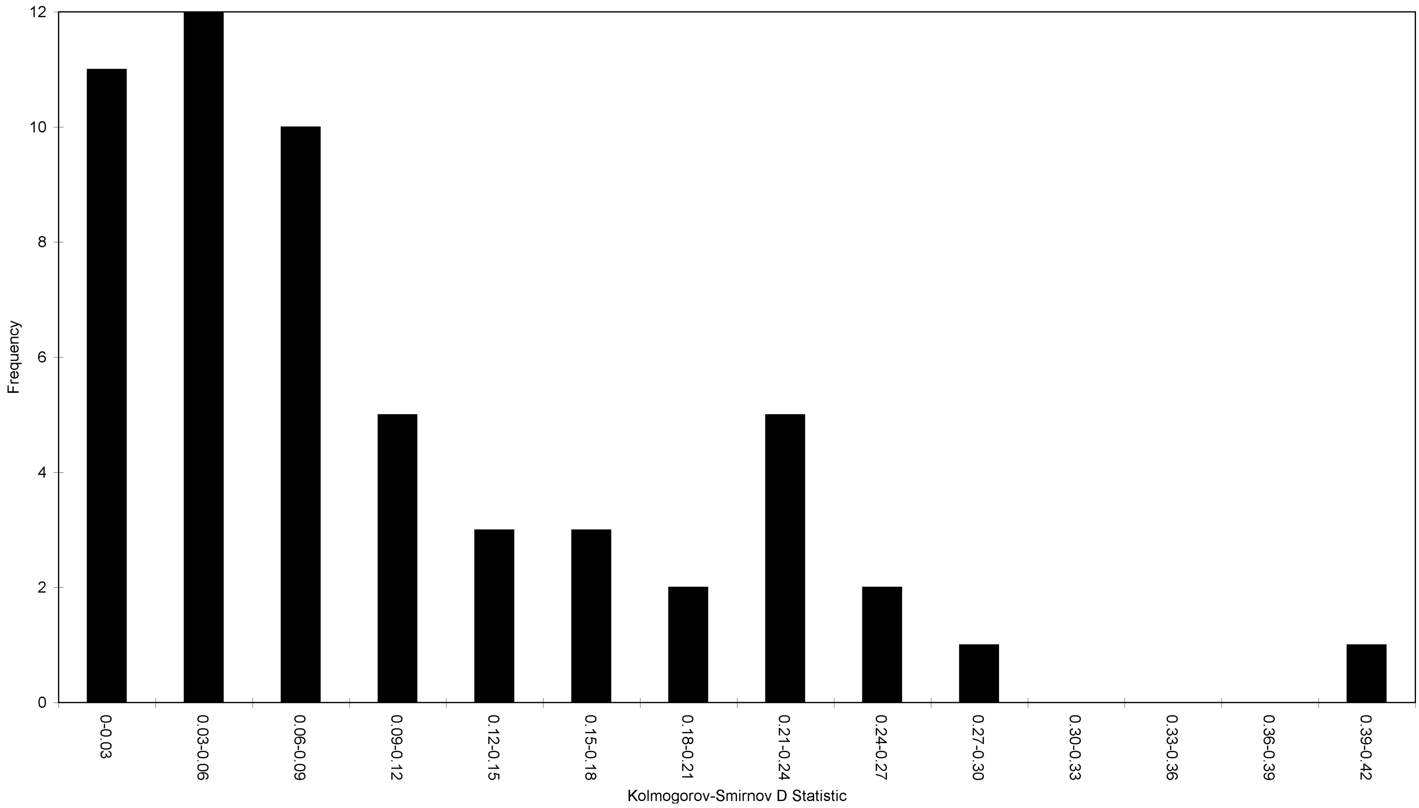
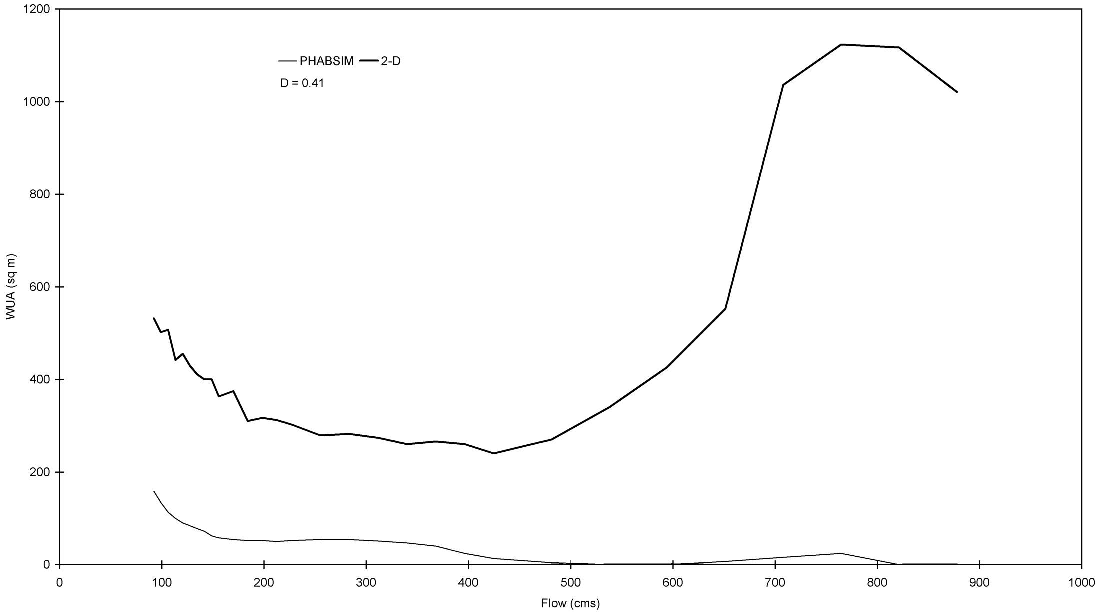


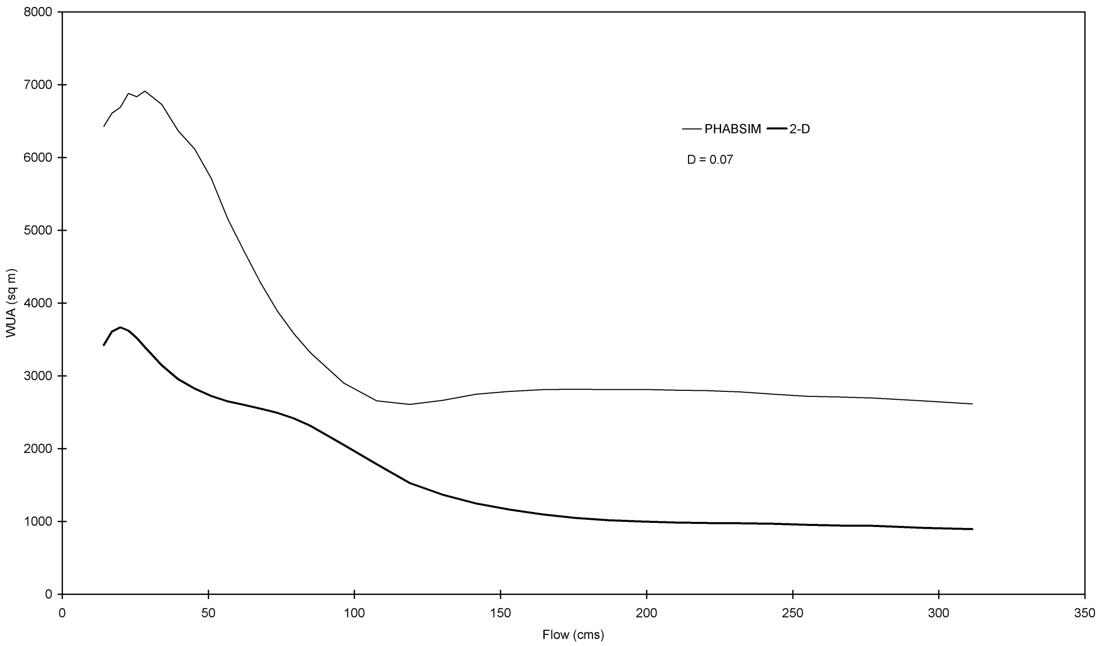
Figure 9 Results of Kolmogorov-Smirnov tests of PHABSIM versus River2D flow-habitat relationships. One of 55 tests was significant at $p = 0.05$.

from 22 with the original channel index file (Figure 11A) to 13 with the test channel index file (Figure 11B). The distribution of flow across the inflow boundary can have a substantial effect on the velocities predicted by River2D, at least in the upper portions of the sites. Accordingly, the performance of River2D could be improved by having a bed topography at the inflow boundary that is proportional to the measured distribution of velocities at the top of the site, so that the thalweg at the inflow boundary would be directly upstream of the highest velocity at the top of the site. The performance of the River2D model could also have been improved by collecting two additional types of data: the bed topography in one channel-width upstream of the top of the site, and mapping polygons of the substrate distribution. The velocity simulation within the site would have been improved by incorporating the actual bed topography upstream of the site into the computational mesh, instead of using an artificial upstream extension, as was done in this study. Since the substrate at a given point is assigned based on the closest point where substrate data was collected, River2D assumes that the substrate changes half-way in between two sets of cross-sectional points. Mapping substrate polygons would more accurately define where changes in substrate occur, and thus improve the performance of River2D with respect to substrate distribution.

The purpose of this study was to compare the habitat predictions of PHABSIM and River2D, rather than to validate either the HSC curves or the hydraulic modeling of PHABSIM and River2D. The performance of PHABSIM and River2D in predicting the CSI of occupied locations should be viewed as a combination of errors due to the predictive accuracy of the HSC curves used and the accuracy of PHABSIM and River2D to predict the depth, velocity and substrate spatial distribution within the sites. The combined errors were tested against fish data (redd



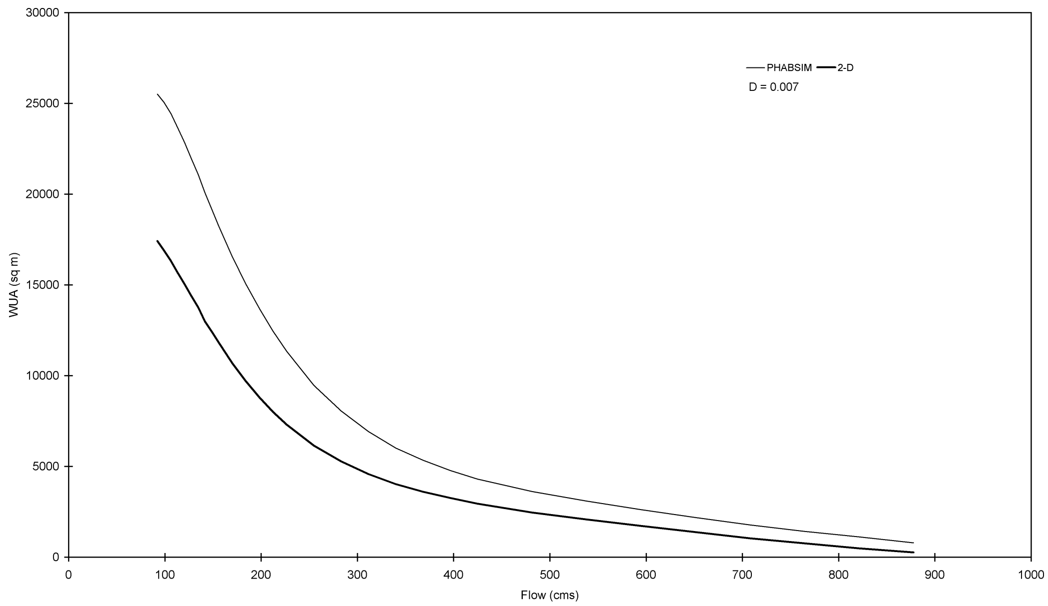
(a)



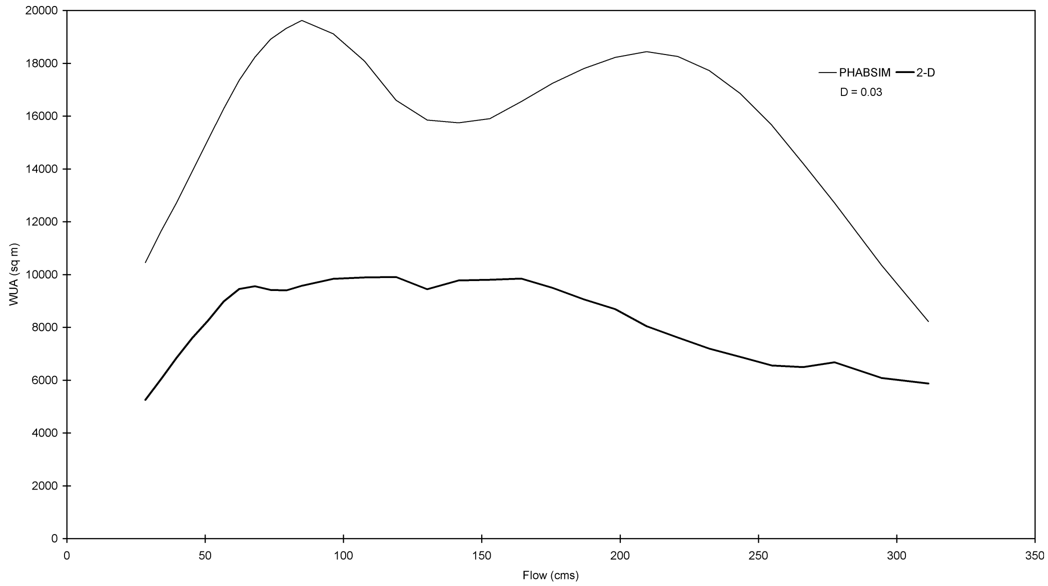
(b)

Figure 10 Sample PHABSIM and River2D flow-habitat relationships. A. Lower Lake Redding (Sacramento River) site, ACID boards out, steelhead spawning. Flow-habitat relationship with highest Kolmorogorov-Smirnov D statistic, $p < 0.05$. B. El Manto (American River) site, fall-run chinook salmon spawning. Flow-habitat relationship with median Kolmorogorov-Smirnov D statistic, $p > 0.05$. C. Upper Lake Redding (Sacramento River) site, ACID boards out, late-fall-run chinook salmon spawning. Flow-habitat relationship with lowest Kolmorogorov-Smirnov D statistic, $p > 0.05$. D. Sailor Bar (American River) site, fall-run Chinook salmon spawning.

locations) across systems and flow levels. Since the same HSC were used for PHABSIM and River2D, differences between the two models in predicting the CSI of occupied locations is entirely due to the ability of the two models to predict depths, velocities and substrates, which are translated into CSI by the HSC. Within the usual use of calibration, the only data used to calibrate the two models were water surface elevations. The data used to develop the HSC (Table 3) could also be viewed as calibration data. Since the redd location data used to compare the habitat predictions of PHABSIM and River2D for the Sacramento and Merced Rivers were a subset of the data used to develop the HSC for these rivers, these data can not properly be considered validation data. In contrast, the redd location data for the American River were not used to develop the American River HSC, and thus the results of the comparisons of the CSI predictions of PHABSIM and River2D can be viewed as a validation of the combination of the American River HSC and the hydraulic modeling of PHABSIM and River2D. The results for each model help to validate the hydraulic modeling of the other, while the combined results of the two models help to validate the HSC.



(c)

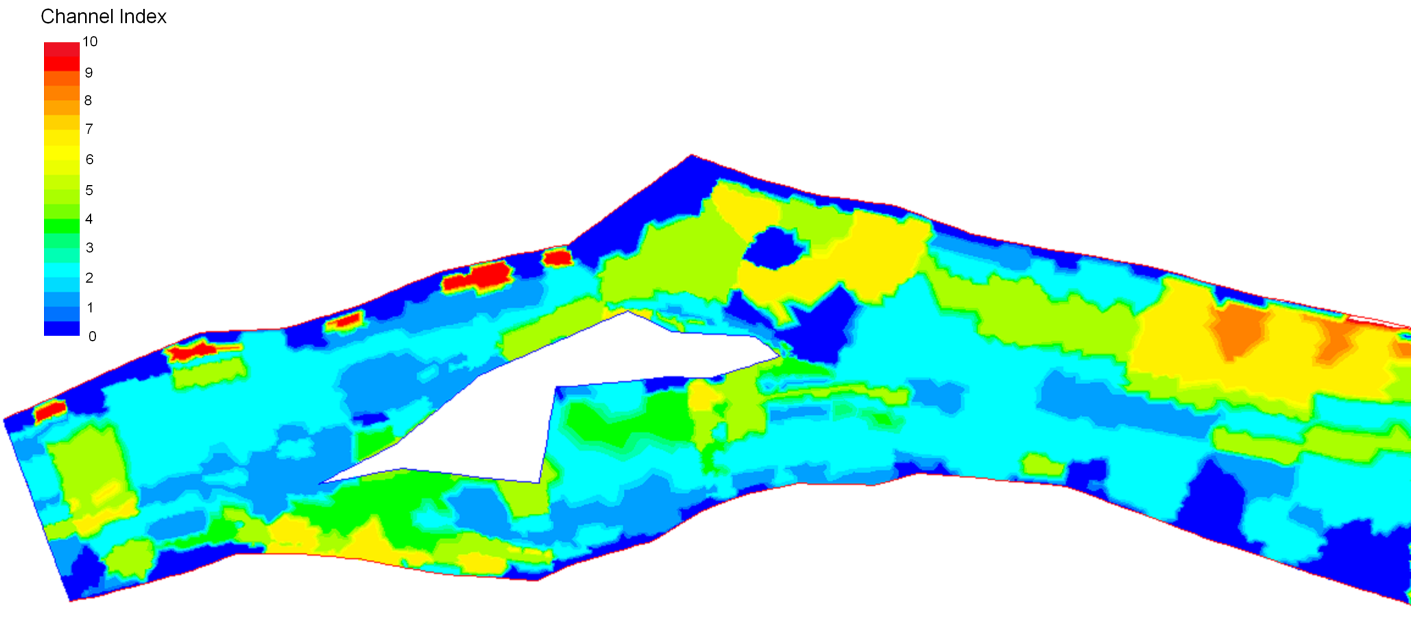


(d)

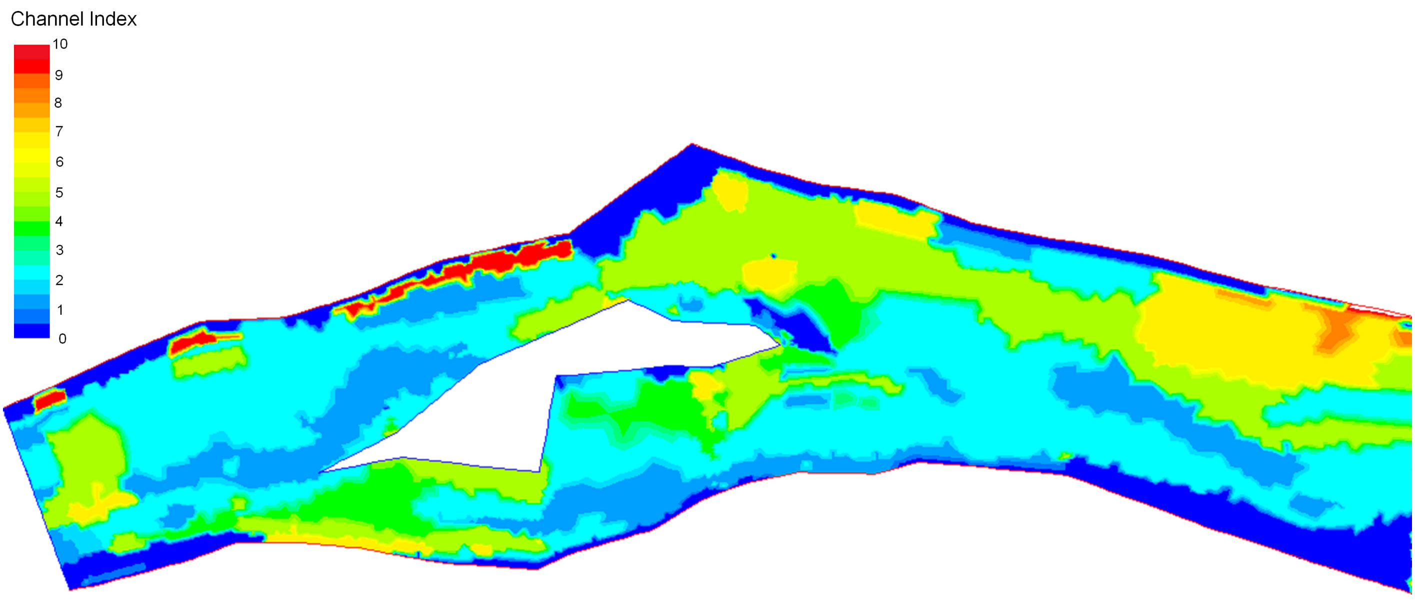
Figure 10 (*Continued*)

There were several limitations of the tests used in this study. The low number of occupied late-fall spawning locations (22 and 16 for, respectively PHABSIM and River2D) resulted in a low power of the Mann-Whitney U test for this race. In this regard, Thomas and Bovee (1993) found in the analagous transferability test that the power of the test was significantly reduced if the number of occupied locations was less than 45. Guay *et al.* (2000) found a significant positive relationship between fish densities and habitat quality indices, similar to our results that the CSI predicted by River2D of occupied locations was greater than for unoccupied locations for the remaining tests. The main limitation of the comparison of the PHABSIM and River2D flow-habitat relationships was that we were not able to compare the flow-habitat relationships of PHABSIM and River2D for areas which could not be modeled with PHABSIM. Similar to the results of this study, Waddle *et al.*, (2000) found mixed results in PHABSIM and River2D's abilities to predict velocities.

This study had mixed results on whether River2D is better than PHABSIM at predicting spawning habitat, and found little difference between PHABSIM and River2D in flow-habitat relationships. However, with the refinements suggested above, River2D has the potential to significantly outperform PHABSIM at predicting spawning habitat. Probably the main advantage of River2D is its ability to model conditions, such as transverse flow, across-channel variations in water surface elevation, and flow contractions/expansions, which cannot be modeled with PHABSIM. If flow-habitat relationships for areas that cannot be modeled with PHABSIM are significantly different from areas,



(a)



(b)

Figure 11 Distribution of substrate predicted by River2D for the American River site in Figures 6 and 8. A. Distribution of substrate using the original channel index file. B. Distribution of substrate using the test channel index file where substrate was determined based on the closest longitudinal substrate datapoint.

such as those used in this study, which can be modeled with PHABSIM, the choice of model would have an effect on instream flow prescriptions.

Acknowledgements

I thank E. Ballard, S. Gallagher, J. Thomas, W. Amy, B. Cordone, J. Kelly, J. Kent, R. Williams, E. Sauls, J. Big Eagle, J. Ly, S. Boring, and K. Brown for assistance with the studies discussed in this paper. I also thank T. Waddle of the U.S. Geological Survey and P. Steffler of the University of Alberta for technical assistance with the River2D modeling. The studies were funded by the Central Valley Project Improvement Act (Title XXXIV of P.L. 102-575).

References

1. Bovee, K.D. (1982). "A Guide to Stream Habitat Analysis Using the Instream Flow Incremental Methodology," *Instream Flow Information Paper*, No. 12, FWS-OBS-82-26, U.S. Fish and Wildlife Service, Fort Collins, Colorado.
2. Gallagher, S.P. and Gard, M.F. (1999). "Relationship Between Chinook Salmon (*Oncorhynchus tschawytscha*) Redd Densities and PHABSIM-Predicted Habitat in the Merced and Lower American Rivers, California," *Canadian Journal of Fisheries and Aquatid Sciences*, 56, 570–577.
3. Gard, M. (1998). "Technique for Adjusting Spawning Depth Habitat Utilization Curves for Availability," *Rivers*, 6, 94–102.

4. GARD, M. and BALLARD, E. (2003). "Applications of New Technologies to Instream Flow Studies in Large Rivers," *North American Journal of Fisheries Management*, 23, 1114–1125.
5. GHANEM, A., STEFFLER, P., HICKS, F. and KATOPODIS, C. (1995). "Two-Dimensional Modeling of Flow in Aquatic Habitats," *Water Resources Engineering Report* 95-S1, Department of Civil Engineering, University of Alberta, Edmonton, Alberta. March 1995.
6. GHANEM, A., STEFFLER, P., HICKS, F. and KATOPODIS, C. (1996). "Two-Dimensional Hydraulic Simulation of Physical Habitat Conditions in Flowing Streams," *Regulated Rivers: Research and Management*, 12, 185–200.
7. GUAY, J.C., BOISCLAIR, D., RIOUZ, D., LECLERC, M., LAPOINTE, M. and LEGENDRE, P. (2000). "Development and Validation of Numerical Habitat Models for Juveniles of Atlantic Salmon (*Salmo Salar*)," *Canadian Journal of Fisheries and Aquatid Sciences*, 57, 2065–2075.
8. LECLERC, M., BOUDREAULT, A., BECHARA, J.A. and CORFA, G. (1995). "Two-Dimensional Hydrodynamic Modeling: A Neglected Tool in the Instream Flow Incremental Methodology," *Transactions of the American Fisheries Society*, 124, 645–662.
9. MILHOUS, R.T., UPDIKE, M.A. and SCHNEIDER, D.M. (1989). "Physical Habitat Simulation System Reference Manual – Version II," *Instream Flow Information Paper*, No. 26, Biological Report 89[16], U.S. Fish and Wildlife Service, Fort Collins, Colorado.
10. STEEL, R.G.D. and TORRIE, J.H. (1980). *Principle and Procedures of Statistics: A Biometric Approach*, McGraw-Hill, New York.
11. STEFFLER, P. and BLACKBURN, J. (2002). "River2D: Two-Dimensional Depth Averaged Model of River Hydrodynamics and Fish Habitat," *Introduction to Depth Averaged Modeling and user's Manual*, University of Alberta, Edmonton, Alberta. http://bertram.civil.ualberta.ca/download.htm
12. THOMAS, J.A. and BOVEE, K.D. (1993). "Application and Testing of a Procedure to Evaluate Transferability of Habitat Suitability Criteria," *Regulated Rivers: Research and Management*, 8, 285–294.
13. WADDLE, T., STEFFLER, P., GHANEM, A., KATOPODIS, C. and LOCKE, A. (2000). "Comparison of One and Two-Dimensional Open Channel Flow Models for a Small Habitat Stream," *Rivers*, 7, 205–220.
14. WILKINSON, L. (1990). SYSTAT: The System for Statistics, SYSTAT, Inc., Evanston, Illinois.
15. WILLIAMS, J.G. (1996). "Lost in Space: Minimum Confidence Intervals for Idealized PHABSIM Studies," *Transactions of the American Fisheries Society*, 125, 458–465.
16. YALIN, M.S. (1977). *Mechanics of Sediment Transport*, Pergamon Press, New York.